Estimating the Value of Offsite Data to Advertisers on Meta

Nils Wernerfelt, Anna Tuchman, Bradley T. Shapiro, Robert Moakler*

July 29, 2022

Preliminary Draft

Abstract

We study the extent to which advertisers benefit from data that are shared across applications. These types of data are viewed as highly valuable for digital advertisers today. Meanwhile, product changes and privacy regulation threaten the ability of advertisers to use such data. We focus on one of the most common ways advertisers use offsite data and run a large-scale study with hundreds of thousands of advertisers on Meta. Within campaigns, we experimentally estimate both the effectiveness of advertising under business as usual, which uses offsite data, as well as how that would change under a loss of offsite data. Using recently developed deconvolution techniques, we flexibly estimate the underlying distribution of treatment effects across our sample. We find a median cost per incremental customer using business as usual targeting techniques of \$43.88 that under the median loss in effectiveness would rise to \$60.19, a 37% increase. Similarly, analyzing purchasing behavior six months after our experiment, ads delivered with offsite data generate substantially more long-term customers per dollar, with a comparable delta in costs. Further, there is evidence that small scale advertisers and those in CPG, Retail, and E-commerce are especially affected. Taken together, our results suggest a substantial benefit of offsite data across a wide range of advertisers, an important input into policy in this space.

1 Introduction

Digital advertising is increasingly popular and constitutes a majority of total advertising spending (Cramer-Flood 2021). From an advertiser's perspective, one of the primary benefits of digital over other channels is the ability to use detailed user data to match ads to consumers. Amongst these data, the most valuable for advertisers is thought to be data such

*Wernerfelt is the first author and authors are listed in reverse alphabetical order. Wernerfelt, Moakler: Meta Platforms, 1 Hacker Way, Menlo Park, CA 94025. Tuchman: Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208. Shapiro: University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637.

as browsing history, past purchase events, or items that are currently in a user's shopping cart online. In this paper, we estimate the value of such "offsite" data using a large-scale experiment across more than a hundred thousand advertising accounts on Meta.

While such data may help target ads better, product and regulatory changes loom that may affect the ability of this information to be shared with advertising platforms. In Europe, the General Data Protection Regulation (GDPR) requires explicit consent for users' individual behavior data to be used for ad targeting. On the product side, Apple's roll out of their "Ask App Not to Track" feature in iOS 14.5 made it easier for users to opt out of data sharing across mobile apps. In the immediate aftermath of that roll out, major advertising platforms lost more than $140 billion in collective valuation (Hackett and Harty 2021). Additionally, there is prospective legislation around the world that similarly would limit data sharing (e.g., updated versions of the GDPR in India, Brazil; the ePrivacy initiative in Europe).

The stated motivations for these policies typically revolve around giving users more privacy. While many users value their privacy, increasing privacy in this manner may come at a cost to advertisers in the form of reduced advertising effectiveness and to users in the form of being served less relevant ads. As ad targeting is a primary use of these data, any holistic assessment of costs and benefits should include the effects of policies on the advertising market.

Apart from understanding the effect of privacy regulation on advertisers overall, effects on small businesses in particular may be especially important and could affect market competition. Small businesses rely heavily on digital and social media advertising, whereas larger businesses generally advertise on more traditional channels (Moorman 2022, Peck 2022). Some segments, such as Direct-to-Consumer, advertise almost exclusively on digital and rely heavily on social media advertising as they grow (Geyser 2020). If digital advertising becomes substantially less effective, it may thus have implications for the competitive landscape across many markets.

In this paper, we estimate two main quantities across over more than a hundred thousand advertisers. First, we estimate the effectiveness of advertising campaigns run on Meta that use ad delivery methods that include offsite data (i.e., data originating off of Meta's sites and applications). We will call this "business as usual" delivery (BAU). Second, we estimate the loss in advertising effectiveness when advertisers lose access to offsite data. We will call this "signal loss." Advertisers use offsite data in a number of ways in practice; hence, we focus on a single, main use case that is the most common on Meta and study effects from that alone. Meta's main source of offsite data comes from digital "pixels" which advertisers place on their websites and from the "Facebook SDK" which is used in mobile applications.[1] Both of these technologies provide information directly to Meta about the behavior people took on websites and apps that help improve ad targeting, even though such data would otherwise be unobserved by Meta since they occur off of Meta's platform. Specifically, we take ads whose delivery is being optimized for sales (an offsite signal) and compare outcomes when the ads are only optimized for link clicks (an on platform outcome). Intuitively this

[1] We describe pixels and SDK's in more detail later on, but two useful background resources are https://www.facebook.com/business/learn/facebook-ads-pixel and https://developers.facebook.com/docs/app-events/overview. Note that nearly all major advertising platforms have their own analogues of these two.

approach constitutes a comparison between the major usage of offsite data and a substitute that only relies on onsite data.

Studying this question comes with the usual difficulty that advertising is, in general, not randomly assigned, so making inferences about advertising effectiveness requires us to isolate random variation. It also comes with the further challenge that advertisers which currently do not employ delivery that uses offsite data are different from advertisers that do in ways that are likely correlated with sales. As a result, we need two dimensions of randomization: (1) within delivery procedure, we must randomize advertising exposure across individuals and (2) within advertisers, we must randomize the ad delivery algorithm (with and without offsite data).

At a high level, we do this by running experiments on ad traffic across a large number of potential purchase events advertisers are trying to drive; specifically, we take all traffic that is optimizing for purchase events and then run two separate experiments on small fractions of that live traffic. First, we randomly hold some users out from seeing the ads, allowing us to estimate ad effectiveness at baseline for the campaigns that are using offsite data. Second, we change a small fraction of their traffic to be delivered as if it did not have offsite data; similarly adding a holdout onto that group allows us to estimate how effective the same campaign is under a loss of offsite data.[2] Repeating this process across hundreds of thousands of products, we are able to make statements about both ad effectiveness at baseline and how much less effective the same campaigns would be without offsite data. Our core outcome variable of interest is the number of incremental customers per dollar, which we sometimes invert to report as the number of dollars per incremental customer.[3,4]

The empirical distributions of treatment effects across experiments are highly skewed and fat-tailed. Further, most experiments are individually too underpowered to make conclusions in isolation. To accommodate these aspects of our data and flexibly estimate the latent distribution of effects, we use the empirical Bayes deconvolution method of Efron (2016). We show how more standard meta-analytic approaches – such as fixed effects or normal distribution random effects – could lead to substantial bias. We describe the procedure in depth in later sections.

Under business as usual targeting using offsite data, we estimate a median cost per incremental customer of \$43.88, with 10th and 90th percentiles \$5.03 and \$172.77.[5] This estimate is based on a global sample of advertisers, similar in spirit to Shapiro et al. (2021). As a result, we view this as a very general estimate of how effective advertising that uses offsite data at Meta is. These numbers are also ballpark consistent with many customer

[2] We note these two holdout groups are needed since the delivery processes endogenously select different users to show the ads to. The respective holdout groups allow us to estimate the effect of the ads conditional on the selected populations.

[3] We focus on customers instead of purchases since the latter is less clean in our data. As we detail later, advertisers install their own pixels, and advertisers may either mis-install the pixel or deliberately configure it in complex ways for reasons not clear to Meta. The empirical distribution of purchases per customer is very skewed, and instead of imposing assumptions on it, we instead focus on the extensive margin, which is much more stable.

[4] In our sample we also do not observe prices of the products sold or total investment by the advertiser (e.g., creative costs).

[5] Note since our estimation procedure allows us to recover the entire underlying distribution of effects, these percentiles refer to points across that distribution; they do not constitute a confidence interval. Also, \$43.88 corresponds to 0.0228 incremental converters per dollar.

acquisition cost benchmarks (though those are often estimated quite differently and on more restricted samples). Finally, we note that given only a small minority of advertisers use measurement tools, these numbers include a large population of advertisers who individually are too underpowered, cannot afford, or otherwise are unable to generate their own point estimates.

Next, we rerun our main specification on the within-product difference in ad effectiveness under a loss of offsite data. Here we estimate a median change in the number of incremental converters per dollar of -0.0062; as comparison, at the aforementioned median BAU estimate, this would change the cost per incremental customer from $1/0.0228$ to $1/(0.0228 - 0.0062)$, or \$60.19. This represents a 37% increase in costs of acquiring new customers. Further, about 90% of the estimated underlying effects lie below zero, suggesting a large share of the advertisers in our final sample will see a decrease in ad effectiveness under signal loss.[6]

We rerun our core models on various demographic subsets to explore heterogeneity. We find evidence that small scale advertisers – which constitute the vast majority of our sample and online advertisers – have effects similar to our main overall estimates; larger scale advertisers, however, are hurt less. At every point in the distributions, small scale advertisers have more effective ads, but they also tend to lose more without offsite data; this speaks to the relative importance of offsite data for smaller advertisers. We also explore effects by the main verticals in our sample (Consumer Packaged Goods, Retail, E-commerce), and similarly find evidence that they will be hurt more than other verticals.

In a final analysis, we look at the purchasing behavior of our users six months after the study was run. Using the randomization induced by our week-long experiment, we can see whether ads delivered with or without offsite data generate more longer term customers of those products. We find evidence that purchase-optimized ads generate substantially more longer term customers per dollar than click-optimized ads, with a magnitude similar to that of the short run effect. This is consistent with the purchase-optimized ads both being more effective at finding more valuable customers for advertisers, as well as delivering a larger benefit on the consumer side.

Our results come with some caveats. First, our experiment is intentionally designed to be partial equilibrium. We are interested in holding all else constant today, what is the value of offsite data to targeting. That is, if advertisers could only use link clicks instead of offsite conversions to optimize their ads, how much less effective would they be? This does not constitute a full scale policy evaluation of a candidate regulation. In particular, it does not capture general equilibrium effects that might come from such a regulation (e.g., R&D into new targeting procedures, ad price changes, signal loss on other platforms, endogenous entry/exit of advertisers or platforms). Instead, we provide a clean, experimental identification of partial equilibrium results on the value of offsite data. Second, and relatedly, our results do not indicate whether using offsite data increases social welfare. While our results are clearly relevant to social welfare, they are not sufficient given that we do not measure profits to the platform, valuations of privacy to consumers or the welfare benefits to consumers of more relevant ads. Given the scope of the paper, we leave the other pieces

[6] As we discuss later, this result is conditional on our final sample, which consists of advertisers who are setting up their campaigns correctly as per Meta's recommendations for using offsite data; this result is does not apply to the unconditional population of all advertisers who are using offsite data. This result also does not speak to whether the shift in ad effectiveness is economically meaningful for any particular firm.

of the social welfare question to future research.

This paper adds to the literature in two main ways. First, we provide the largest, most credible and most generalizable study of the effects of targeted advertising on Meta with offsite data. Our sample is the near universe of advertisers on Meta which use offsite data to optimize targeting for conversions. Ultimately, this is over 100,000 advertisers, and we have randomized advertising exposure, ensuring credible and robust estimates. Previous research has studied the effectiveness of digital advertising using field experiments (Goldfarb and Tucker 2011a, Lewis and Rao 2015, Johnson et al. 2017, Gordon et al. 2019), but largely has not focused on the value of behavioral targeting. Other work has focused on behavioral targeting, but at a much smaller scale, not always with experiments, and not designed specifically to avoid selection into the sample. (Bleier and Eisenbeiss 2015, Sahni et al. 2019, Rafieian and Yoganarasimhan 2021, Johnson et al. 2022). For context, perhaps the most similar study in spirit to ours (large scale, across advertiser study, designed to have no selection into the sample on outcome) is conducted an offline setting is Shapiro et al. (2021), which includes 288 advertisers, all in CPG, and uses observational methods to isolate random variation. We view this as a major managerially relevant contribution in its own right, as it helps potential advertisers on Meta generate an unbiased prior distribution of how well advertising might work for them to acquire customers.

Second, we provide the most comprehensive and generalizable evidence to date on the value of offsite data for ad targeting on a major platform. This is useful in two ways. First, it helps advertisers understand how much less they should be willing to pay for advertising that cannot use offsite data. Second, it is a useful input to policy makers hoping to conduct a holistic policy evaluation of potential privacy regulations that would restrict the use of offsite data. Rafieian and Yoganarasimhan (2021) is similar to this study in that it considers a tradeoff between the sensitivity of data used in ad delivery and outcomes for different actors in the advertising ecosystem. Importantly, our paper distinguishes itself by measuring actual sales rather than click through rates as the main outcome of interest, by using a named large ad platform (Meta), and by using experimental variation across over 100,000 advertisers to estimate effects.

Other related papers include those that study the effects of privacy regulations that effectively limit firms' ability to run targeted advertising campaigns. Many of these papers focus on non-sales outcomes such as equilibrium advertising prices (Johnson 2013) and investment in technology firms (Jia et al. 2018). Within the privacy literature, Goldfarb and Tucker (2011b) and Goldberg et al. (2021) are most closely related to our work. Goldfarb and Tucker (2011b) exploits quasi-experimental variation in the adoption of advertising regulations in the EU to compare survey responses to ads in EU countries in which ad targeting was limited to non-EU countries where richer targeting was still permitted. Goldberg et al. (2021) also leverages quasi-experimental variation in the implementation of privacy regulations over time to evaluate the impact of coarsening firms' information sets on advertising effectiveness.

Compared to the extant literature, we believe the two core advantages of our paper are the experimental variation combined with our sample. In addition, to the best of our knowledge, we are the first to apply the deconvolution methods of Efron (2016) in a marketing setting, thereby introducing a powerful new tool to this literature. On our sample, generating experimental estimates for campaigns of both their effectiveness today and how much that would change under a counterfactual world essentially requires internal access at a major advertising platform – as well as access to substantial engineering resources. We were able to solve this difficult barrier to research on this topic. Further, Meta is global platform, with more than 3.6 billion monthly active users and more than 10 million advertisers; as we describe in our sample recruitment process, we are able to experiment on a large sample of advertisers with minimal selection concerns. While we focus on a usage of offsite data that happens across major advertising platforms, the sample itself is economically meaningful per se.

The organization of this paper is as follows: Section 2 describes more background on digital advertising and the relevant context of our paper; Section 3 walks through the experimental design; Section 4 describes our sample; Section 5 contains an example to build intuition for our analysis; Section 6 has the main results; and Section 7 concludes.

2 Background

We start by providing some context on digital advertising and offsite data that will prove useful in understanding and interpreting our results.

Digital advertising relies on a large set of signals and data as well as complex modeling processes to deliver ads effectively for advertisers and consumers. One way to categorize the signals and data used in online ad delivery is to bucket them into onsite or offsite data. The exact definitions of these two groups will vary based on who is delivering ads. For example, at Meta, data collected on the platform such as a user's interaction with the Facebook or Instagram applications, the number of comments they write, or the Pages they Like can be considered onsite data. Onsite signals like these are used in matching users and ads. On the other hand, offsite data are also used in delivery and traditionally are viewed as more valuable for the matching process. Offsite data, as its name suggests is data that originated from off of the platform utilizing it. Again, considering what this looks like for Meta, offsite data typically comes from sources such as the "pixel" and "Facebook SDK."

Many digital advertising platforms (Meta included) can set up pixels on their websites or implement SDKs on their mobile applications to help with ad delivery (both in terms of measuring an ad's effectiveness as well as optimizing ad delivery to consumers).[7] Pixels are a few lines of JavaScript code that an advertiser can install on their website. When a person visits the website or performs certain actions on it (e.g., purchasing a product), the pixel is "fired" and some data is transmitted back to Meta. Similarly, as pixels are used on desktop websites, solutions such as Facebook's SDK are used to perform a similar action in mobile applications. When an app is installed or a particular action is taken in an app with the SDK integrated, the SDK can report certain pieces of information about the event and user. In this paper, we refer to "offsite data" as data coming from the pixel and SDK given they come from outside of Meta.[8] Note we restrict our definition of offsite data such that it does not include data that advertisers may import into Meta from other means (e.g., uploading

[7] The common pop-ups that appear on many websites asking for consent to use cookies are in part to cover usage of pixels.

[8] Similarly, many large digital companies such as Google, Twitter, Snap, and TikTok have their own pixels and SDKs that collect and send data to them to help with ad delivery and measurement.

a list of email addresses to target or demographics of users that may have come from a data vendor or other source). It is worth highlighting that while these data are offsite from the view point of Meta, the data are *onsite* for the advertisers who are running ad campaigns. When we discuss data such as these, we view them from the lens of a platform delivering ads on behalf of advertisers which means data on "outcome" events such as purchases will almost exclusively be categorized as offsite.

As we have already mentioned, offsite data are used in many ways in online advertising. We do not aim to articulate every use case and measure the corresponding effects; rather, we focus on arguably the main way advertisers use such data across platforms today, and estimate what would happen if they instead could only rely on onsite data.

Specifically, our focus is on the usage of offsite data in optimizing ad delivery. Nearly every major online platform enables advertisers to optimize their ad delivery according to different objectives or outcomes, and such optimization is viewed as one of the key advantages of online advertising over traditional channels. In traditional channels, dynamic updating of an advertiser's target audience is often constrained to be a slow process. For example, an advertiser may run a television ad, estimate its effect on an outcome of interest, update the target audience, and then repeat. In contrast, with online advertising, it is much easier to get real time feedback and adjust accordingly. This ability to utilize real time granular data is the heart of digital ad optimization and is one of its most substantive advantages.

One of the reasons the distinction between onsite and offsite data is so critical is because many of the objectives that advertisers care most about happen off the platforms where their ads are shown. With pixel and SDK data, an advertiser can specify to optimize delivery on a platform like Meta not just for onsite actions such as clicks or engagement, but rather for offsite outcomes like purchases or email signups directly. For example, if an advertiser specifies to optimize for purchases, Meta could show the ads to a number of users, see who purchases the product of interest, train a machine learning model to predict who is likely to purchase, show the ads to a new set of users, and retrain or update their delivery model to maximize an ads efficiency.[9]

Therefore, even if an advertiser does not know (or is mistaken about) her target audience, such optimization procedures can help effectively find or fine tune a target audience to show the ads to. Getting ads in front of the relevant audience and minimizing the cost per ad and the cost per outcome from those ads, is a core metric that advertisers optimize for. As the goal of many ad campaigns is to drive consumers to purchase a product or engage with the brand, the ability to optimize towards these specific actions is critical in minimizing costs for the advertiser. However, we make the important note that the optimization procedure outlined here does not optimize for *incremental* conversions. Hence, a criticism some have made around optimization procedures such as these is that they could simply be adept at finding individuals who were already going to purchase the advertised product. How large a concern this should be is an empirical question which we take seriously and use as a core

[9]Two technical notes: (1) the actions advertisers may log with pixels or the SDK can cover many standard "event types" such as Add to Cart, Purchase, or View Content. In selecting an optimization goal that uses pixel data, an advertiser must specify which event type to optimize for; as we will detail later, our experiment focuses specifically on ads that were optimizing for purchases. (2) The actual training and deployment of these delivery models leveraging the data we discuss is more complex than what is described here due to the scale and speed at with digital ad auctions occur at.

tenet to the structure and design of our measurement procedure, as we detail next.

Finally, we note that offsite data has other use cases in the ad delivery funnel other than for optimizing towards specific events. For example, it can be used as inputs into "lookalike audiences" (where an advertiser would specify Meta should show ads to people "similar" to a specific audience list). Hence, we are not considering the complete counterfactual of if advertisers lost all offsite data, but rather focusing on arguably the primary use case (ad delivery through optimization) and looking at effects from that. We discuss more details of this counterfactual later.

3 Experimental Design

For a broad sample of advertisers who are using offsite data to optimize the delivery of their ads, our aim is to estimate both (i) the effectiveness of their ads with their baseline strategy as well as (ii) how much less effective they would be under a loss of offsite data. At a high level, we do this by running experiments on ad traffic across a large number of advertisers that leverage pixels and SDKs for their ad strategies. Specifically, we take all traffic that is optimizing for purchase events from an offsite signal (we will refer to this process as "Offsite Conversion Optimization" going forward even though we focus specifically on purchase conversion events) and then experiment on a small fraction of it to estimate both quantities of interest for each set of ads. This process generates more than a hundred thousand experiments, which we can collectively parse and analyze. We describe our setup in more detail below.

First, consider the set of all ads that are using Offsite Conversion Optimization as a delivery strategy. Our core design is to take a small fraction of that overall traffic and implement two changes. The first change is that we randomly hold a small fraction of users out from being exposed to the campaign. Since the optimization procedure endogenously finds users to show the ads to, treatment and control are determined conditional on selection: users where the ad in question won the auction will either be in test or control, according to a random 90%-10% split.Finally, we note that this is an intent-to-treat analysis, as even though an ad won the auction, a user may not actually see it for a variety of reasons (e.g., scrolling too fast by it, not scrolling far enough down, loss of network connection).

Conditional on assignment to treatment and control, we observe outcome variables for both groups from the offsite sources described in Section 2. This means we can estimate the incremental impact of the campaigns on our purchase outcomes of interest. Incorporating ad spend data, we are able to deduce estimates of the number of incremental purchasers (customers) per dollar spent by advertisers on their campaigns. Note that since these campaigns use Offsite Conversion Optimization, combining all these experiments together in a meta-analysis generates our estimate of the baseline effectiveness of advertising with offsite data.

Our second modification to the live campaigns allows us to estimate the change in incrementality induced by the loss of offsite data. For this, we again take a small fraction of the overall traffic, but here we modify the optimization procedure to not be Offsite Conversion Optimization but rather to approximate a procedure that optimizes for onsite clicks (we will refer to this as "Link Click Optimization" going forward). This optimization mimics a process that finds users who will click on the ad instead of make an actual purchase event. We discuss the justification for this counterfactual and how it was implemented shortly. Again, here we add a user-level randomization component, allowing us to measure incremental effects. Critically, even though we changed the optimization and delivery procedure, we still observe purchase events from the pixels, allowing us to estimate the number of incremental customers per dollar again, albeit when ads are delivered via this modified process.

Put together, this means for each set of ads tied to a purchase event, we observe estimates of the number of incremental customers per dollar under both delivery with offsite data and delivery without. Comparing the within-product changes in treatment effects under another meta-analysis allows us to estimate a combined average effect of loss of offsite data. For a visual display of our experiment, see Figure 1 below.

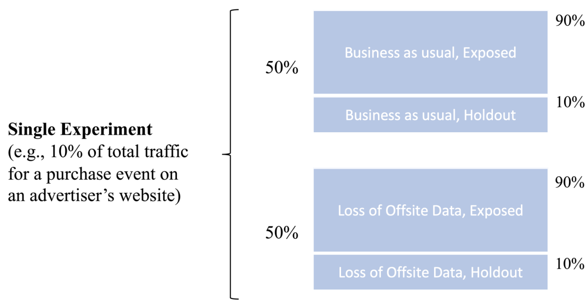


Figure 1: Overview of core experimental design.

Finally, two features we want to emphasize of our experimental design. First, only a small minority of advertisers implement available incrementality measurement tools (such as Conversion Lift[10]); by implementing our measurement component directly onto a small portion of ad delivery, we are able to generate estimates of ad effectiveness for a wide sample of advertisers whose effects would otherwise be unmeasured. Second, and relatedly, this design allows us to scale to large numbers of advertisers in our final sample, allowing us to make statements that are not based on a small number of niche advertisers but are more representative of the overall relevant population.

3.1 Selection of the Link Click Optimization Counterfactual

If advertisers lose the ability to optimize for offsite conversions, what would they substitute to? We make the assumption here that if they could not optimize for offsite events (typically the final outcome of interest for a direct response advertiser), advertisers would instead optimize for the next closest event that could be measured along the "purchase funnel." In the setup of many online ad campaigns this onsite event would generally be an ad click, the outcome farthest down the purchase funnel that is observed on platform.

In addition, empirically the two main optimization types direct response advertisers use are Offsite Conversion Optimization and Link Click Optimization, and one of the most frequent questions advertisers ask Meta is how to choose within this choice set. Given this collective background, we decided to focus on these two optimization methods.

[10] https://www.facebook.com/business/m/one-sheeters/conversion-lift

3.2 Implementation of the Link Click Optimization Counterfactual

At a high level one of the most important aspects of optimization methods are the underlying machine learning models that are used to predict probabilities that a user will take the intended action. For Offsite Conversion Optimization the underlying models will try to predict the probability that a user will perform the conversion event an advertiser is optimizing for (the "conversion rate"). For Link Click Optimization the models predict the probability that a user will click the ad (the "click through rate"). In practice, switching a small share of traffic from live campaigns from using Offsite Conversion Optimization to Link Click Optimization is prohibitively challenging from an engineering perspective. The optimizations and their underlying models are complex and simply swapping them is not possible. However, we instead were able to approximate Link Click Optimization via a two step process.

First, we changed the traffic from Offsite Conversion Optimization to instead be optimized via a model that used recent historical data to predict click through rates that are needed for Link Click Optimization. We then ran a second experiment where we took a sample of campaigns that were using true Link Click Optimization and switched a small amount of their traffic to use the intermediate modeled version of Link Click Optimization we designed. This allows us to scale any potential discrepancies between our intermediate modeled version of Link Click Optimization and in-product Link Click Optimization and therefore compare between performance from Offsite Conversion Optimization and what an advertiser could expect had they moved to Link Click Optimization.

In practice, the intermediate modeled and in-product Link Click Optimization model performed very similarly. The cost per purchase of the modeled version was 1.1 times higher than that for the in-product model, however, and so we scaled the costs from the intermediate model accordingly in our main analysis. Rerunning the main results only focusing on Offsite Conversion Optimization versus the intermediate model yields very similar estimates, as does running on other multiplier values around 1.1.

4 Sample

4.1 Recruitment and Selection

For our sample, we took the universe of advertisers on Meta who used Offsite Conversion Optimization in the three month window before our experiment launched.

We sent a notice to these advertisers through the main surfaces through which advertisers purchase ads on Meta that described the experiment and offered them the chance to opt out. An advertiser was only included in the experiment if they were both exposed to the opt out notice and did not choose to opt out; not being served the notice did not default an advertiser into participating.[11] An example of what the notice looked like can be seen in Figure 2. The

[11] Some interfaces through which advertisers purchase ads are not amenable to notices (e.g., command line APIs). If these advertisers logged into the main interfaces they would have seen the notice, but otherwise, since they would not be exposed to the notice, they were not included in the experiment population. Only a small fraction of all advertisers

notice was translated into 31 of the most frequently used languages Meta's advertisers use and was ultimately sent to four million accounts. The notice would appear each time an advertiser would log in (up to three times) and then would disappear.

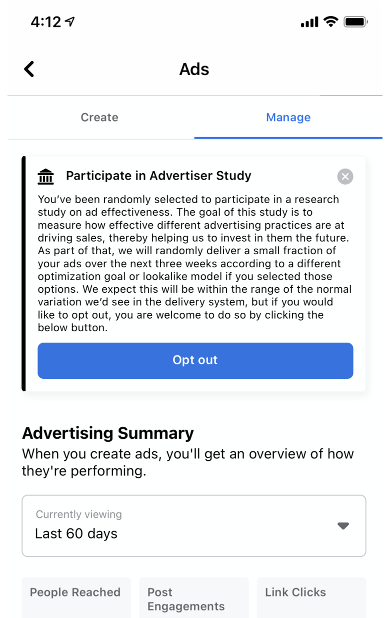


Figure 2: Example illustration of how the recruitment notice would appear to advertisers.

We note that amongst all advertisers who use Offsite Conversion Optimization and were eligible to see the notice, only a subset visited the recruitment interfaces while it was live. Further, amongst those who did not opt out, only a subset of that population ran campaigns again during the experiment that used Offsite Conversion Optimization for purchases. In practice, amongst those who saw the notice, 94% did not opt out, leaving us with 219,664 advertiser accounts who were eligible to be in our experiment. Of those, 187,922 ran ads with Offsite Conversion Optimization for purchases during our experiment.

Of those 187,922, we restrict to experiments that met Meta's minimum recommended number of purchases for Offsite Conversion Optimization to work. Specifically, we filter to include those experiments where the number of purchases in the business as usual exposed condition was greater than or equal to 50 weekly purchases.[12] Conditioning on this minimum actually reduces the sample substantially, leaving us with 70,909 experiments.[13] This

use such surfaces, and, as we show, the demographics of our final sample do match well with the corresponding overall population.

[12]See https://www.facebook.com/business/help/197634954160445?id=561906377587030

[13]We note since each experiment can include ads from potentially multiple accounts, the number of ad accounts in the final sample was 107,114. Such instances may arise when, for example, an advertiser is working with an ad agency who may run ads from a separate account but that still target the same products to sell. In our final sample, 85% of experiments have one account, with less than 1% having more than 10 accounts. When there were multiple accounts associated with an experiment, to define a mapping between experiments and accounts, we took most common demographic within each experiment. For example, if an experiment involved 10 accounts and 8 were from the US, we would label the experiment as from the US. Rerunning under different definitions does not meaningfully affect our results.

reduction means that a large number of accounts were not actually using the product in accordance with recommended best practices, suggesting likely inefficiencies in their advertising. This also means our main reported effects are conditional on the population who are using Offsite Conversion Optimization in accordance with recommended usage, not the population who use the product overall. We focus on the former as these guidelines are publicly available, advertisers can observe if their campaigns are hitting these minimums, and we think that population better speaks to the potential value loss at stake from threats to offsite data. [Further analysis of overall population to be completed.]

4.2 Demographics of Advertisers

Our advertisers came from 161 different countries; as can be seen in Figure 3, our sample has fairly broad representation across regions. Note the log scale in the figure, however: the US is the most represented country, with around 22% of our sample, with other major countries China (7%), Brazil (6%), and India (4%).

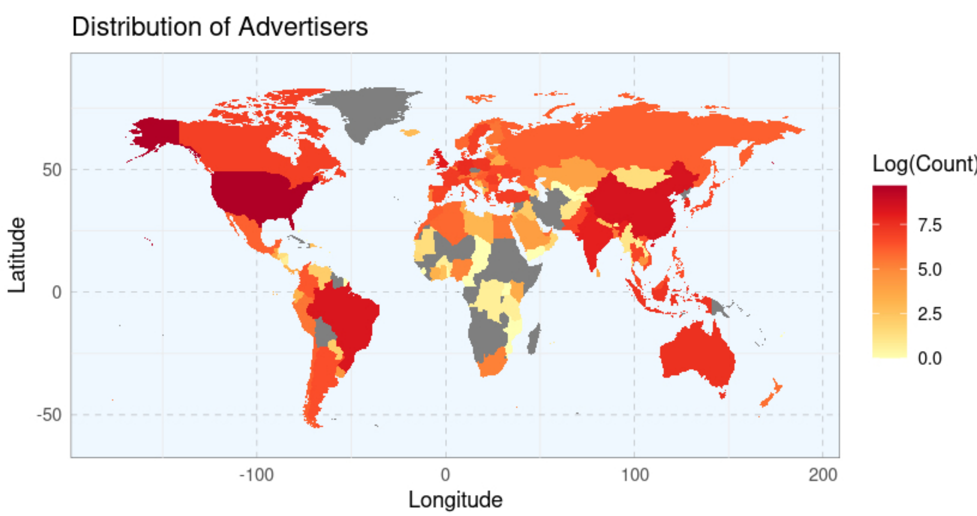
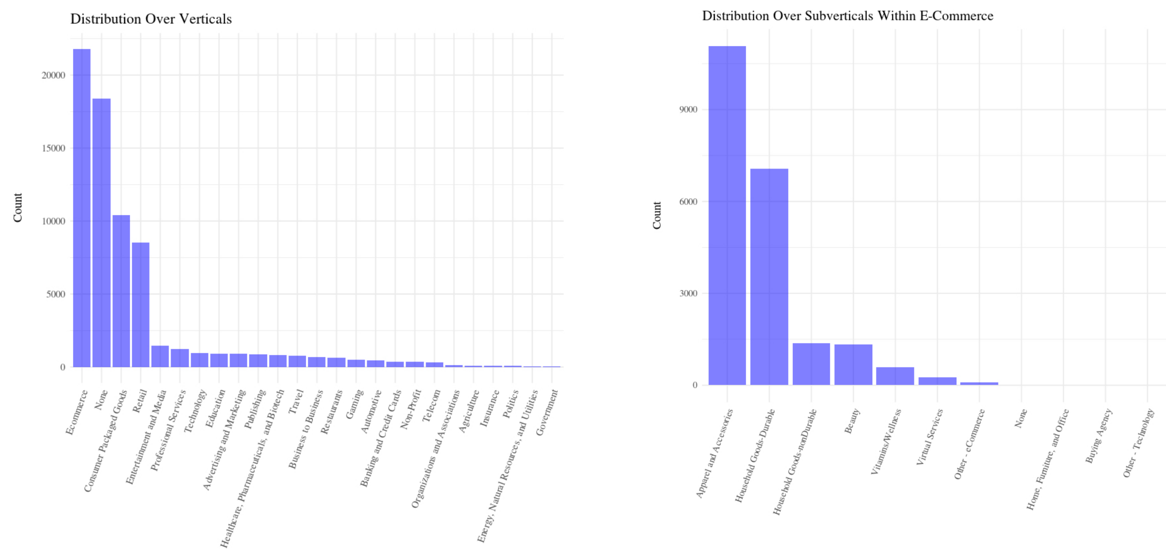


Figure 3: Geographic distribution of advertisers in our sample.

Industry-wise, the most heavily represented verticals are E-commerce (44%), Retail (20%), and CPG (12%), with a long tail after that (Figure 4a). E-commerce has a broad range of subverticals associated with it, so in Figure 4b we show a further, within-vertical breakdown. The classifications of accounts into verticals are based on internal machine learning predictions based on a broad set of data.



(a) Distribution over verticals.

(b) Subvertical representation within E-Commerce.

Figure 4: Distribution of sample over verticals.

As described in our Recruitment and Selection subsection, though we tried to minimize it, there was the possibility for selection bias along the recruitment funnel. To assess the representativeness of our sample, below we show the difference between demographics of our final sample and the demographics of the population of Meta advertisers who were using Offsite Conversion Optimization for purchases with more than the minimum number of recommended purchase events the week of our experiment. Given the size of the populations, many of the demographics are significantly different. However, the magnitudes of the differences are quite small: the average absolute value of the percentage point difference is 1.5% across all demographics despite substantial variation in levels. This makes us feel better about the representativeness of our sample, though the significant differences are a notable caveat.

Finally, an important aspect of our sample is that they are predominantly not spending substantial amounts of money. In the next subsection we report statistics on spend in our sample, but as a preview, Meta has an internal classification of if advertisers are "small" or "large" that is largely determined by ad spend, and by that metric, 87% of our sample were categorized as small. To add more color to that definition, from previous on platform surveys of advertisers we know that ad spend and offline number of employees are significantly negatively correlated and further that conditional on being a "small" advertiser by Meta's internal definition, 90% have fewer than 250 employees (a common definition of a small business), and conditional on being a "large" advertiser, a majority (54%) have greater than 250 employees. Later we will report results broken out by Meta's internal small and large definitions, but at a high level, we can think of these as proxying for on platform spend and as (imperfect) correlates of offline size.

Table 1: Sample representativeness. Three asterisks denotes $p < 0.01$ in a two-sided t-test on the difference. The mean difference corresponds to the mean from the Meta population minus the mean from our experimental sample. Standard errors in parentheses.

	Experimental Sample	**Meta Population**	**Difference**
Vertical			
CPG	0.1196	0.095	-0.0246***
	(1.2e-03)	(0.0017)	(0.0021)
Retail	0.1982	0.2059	0.0077***
	(1.5e-03)	(0.0023)	(0.0028)
E-commerce	0.4443	0.387	-0.0573***
	(1.9e-03)	(0.0028)	(0.0034)
Region			
Asia Pacific	0.2977	0.277	-0.0207***
	(1.7e-03)	(0.0026)	(0.0031)
Europe, Middle East, Africa	0.3673	0.386	0.0187***
	(1.8e-03)	(0.0028)	(0.0033)
Latin America	0.0954	0.0921	-0.0033
	(1.1e-03)	(0.0017)	(0.002)
North America	0.2389	0.2428	0.0039
	(1.6e-03)	(0.0025)	(0.003)
Months since first ad creation			
Less than 1	0.0664	0.0834	0.0169***
	(9e-04)	(0.0016)	(0.0019)
1-6 months	0.157	0.1578	0.0008
	(1.4e-03)	(0.0021)	(0.0025)
6-12 months	0.1228	0.125	0.0022
	(1.2e-03)	(0.0019)	(0.0023)
Greater than 12 months	0.6377	0.6339	-0.0039
	(1.8e-03)	(0.0028)	(0.0033)
Other			
Meta Managed Account	0.3716	0.4037	0.032***
	(1.8e-03)	(0.0028)	(0.0034)
Active Last 7 days	0.9787	0.9994	0.0207***
	(5e-04)	(0.0001)	(0.0006)

4.3 Summary statistics of experiments

Our advertisers are organized into 70,909 experiments that we ultimately use to estimate our distribution of effect sizes. We here present summary statistics of those experiments.

As mentioned in our Experimental Design section, our experiments are created by taking a small percentage of the fraction of traffic from live campaigns. Given the long tail of small spend by advertisers, this means that we have a large number of small experiments – in terms of both total number of users per condition and spend. Below we report summary statistics on these numbers across all our final sample. [Further summary statistics on spend, randomization checks to be added]

Table 2: User count per experiment. Note: 'BAU' refers to business-as-usual (inclusion of offsite data) and 'Test' refers to targeting without that data.

Arm	Min	1st	25th	50th	75th	99th	Max	Mean	SD
BAU Exposed	10	111	2,264	6,911	18,862	244,160	9,295,122	22,352	83,827
BAU Holdout	1	12	251	767	2,095	27,162	1,032,782	2,483	9,314
Signal Loss Exposed	9	112	2,643	8,384	23,438	335,523	9,374,392	28,796	106,136
Signal Loss Holdout	1	12	293	931	2,605	37,336	1,043,232	3,199	11,794

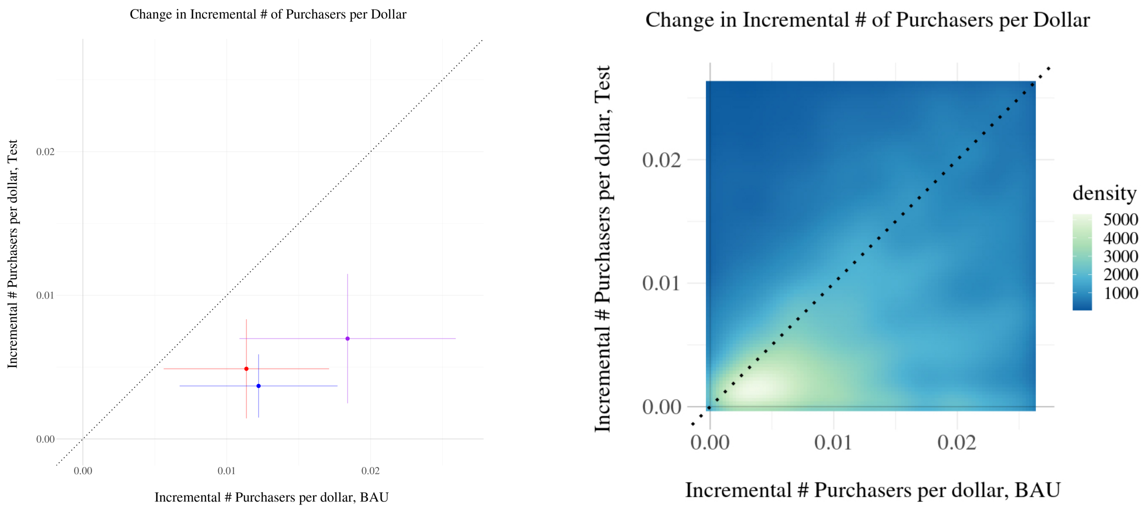
Table 3: Number of converters per experiment. Note the long tail of experiments with very few customers.

Arm	Min	1st	25th	50th	75th	99th	Max	Mean	SD
BAU Exposed	3	3	4	7	16	821	409,681	76	1,977
BAU Holdout	0	0	0	0	2	84	44,273	7	211
Signal Loss Exposed	0	0	2	5	14	782	407,069	72	1,963
Signal Loss Holdout	0	0	0	0	1	79	44,214	7	210

5 Illustrative Example

To set the stage, here we visually show results from (i) three example advertisers, and then (ii) how that graph would look with a heat map of all treatment effect estimates on it. (This is similar in spirit to L'Abbé plots that are often used in meta-analyses.)

In the left panel of Figure 5 we plot the treatment effects and 95% confidence intervals for three example advertisers under inclusion (x-axis) and exclusion (y-axis) of offsite data. The red dot is an apparel company, the blue dot is a beauty company, and the purple dot is a jewelry company.



(a) Example with three SMB advertisers

(b) Heatmap of all experiments.

Figure 5: Motivating graphs.

Note that lying below the 45 degree line means that the advertiser is estimated to get fewer incremental converters per dollar under loss of offsite data (and for each of these specific advertisers we can report out the estimated change in costs). For example, the point estimate for the jewelry company suggests their cost per incremental customer will increase from \$54 to \$154 (derived from the inverse of the number of incremental purchasers per dollar).

Next, on the right, we show a heatmap that includes all the treatment effect estimates in our sample. We can clearly see how there is more density below the 45 degree line: the mass of our estimated treatment effects across all our pixels suggests a negative effect. To estimate how negative of an effect there is, we now turn to our main results.

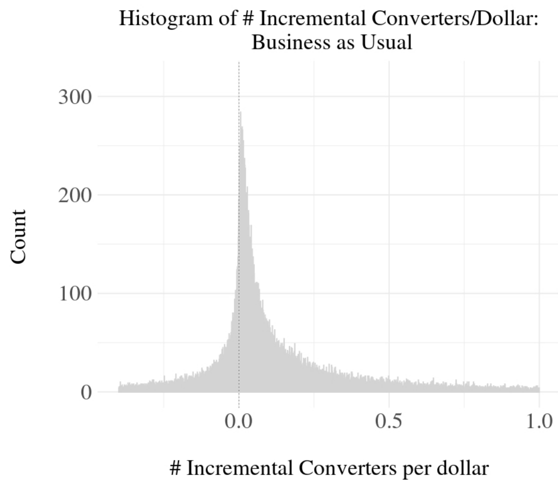
6 Main Results

6.1 Overview

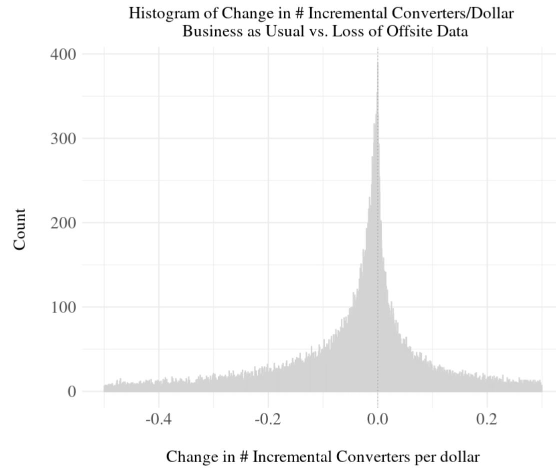
First, to provide a high level overview of our data, below we both plot the distributions of the treatment effects of interest and report out simple summary statistics from those distributions. Our focus here is on the number of incremental converters per dollar under business as usual (inclusion of offsite data) and the within-product change in that number. To estimate each of these numbers, we will run two different meta-analyses in parallel.

Several observations about the distributions below. We note that these observations both help frame the data and motivate the next section, where we conduct our main analysis.

First, note how each distribution is quite skewed. Many meta-analytic methods assume a normal distribution of effect sizes, though for each of our distributions that would introduce substantial bias. As we detail in the next subsection, we thus chose a method that allows a



(a) Business as usual



(b) Within-product difference.

Figure 6: Distribution of observed treatment effects for each of advertising effectiveness today and the within-product difference (bin width of 0.0005).

very flexible functional form to be fit to the data.[14]

Second, there is substantial – and economically meaningful – heterogeneity across experiments. The Table below has summary statistics from these distributions and we can see the 25th and 75th percentiles differ by around 20x and 5x for business as usual and the within-product difference, respectively: for the former, that is the difference between ∼\$100 and ∼\$5 per incremental customer. This suggests instead of estimating a single treatment effect across all experiments, a distribution of effects is more appropriate. Our approach in the next section flexibly accommodates either, but as a preview, we find evidence of non-degeneracy.

Table 4: Summary statistics of distribution of treatment effects for incremental converters per dollar and the within-product difference.

	Min	1st	25th	50th	75th	99th	Max	Mean	SD
# Incremental Converters	-339.025	-3.837	0.010	0.054	0.198	3.997	5,044.898	0.157	13.342
Δ Converters	-1,515.79	-4.499	-0.14	-0.02	0.03	4.81	13,256.55	0.03	34.17

Finally, note how the tail values of these distributions are quite extreme. We are trimming the graphs here at the 5th and 95th percentiles, so the top 5% of treatment effects, for

[14] Another recent meta-analysis with a skewed distribution of effect sizes is DellaVigna and Linos (2022). To accommodate the skew, their estimation procedure uses a mixture of normals fit via maximum likelihood; given the size of our data, we found that approach unstable here and, as detailed later, we further use a method that allows a more flexible functional form.

example, are reporting substantially more than one incremental customer per dollar. While theoretically possible, this stylized fact further motivates aspects of our main analysis for two reasons. First, this distribution does not include information about the variances of the estimates; if the large point estimates are very noisy, we would want to treat them differently than more precisely estimated effects in estimating a combined distribution. Second, we note an institutional detail that advertisers are responsible for installing their own pixels and SDKs, and empirically, they do so with all sorts of noise.

To expand on this point, there are advertisers both who make mistakes in installing their pixels/SDKs and advertisers who try to game Meta's ad delivery platform by specifically setting up them up in nonstandard ways. For a historical example, some advertisers have incorrectly thought if their pixels register a substantial number of purchase events that Meta may give them preferential treatment in distribution; in line with this, some advertisers set up pixels so each web page view would be set to Meta as a purchase event. Alternatively, some advertisers accidentally may register the wrong pixel or SDK to a campaign (e.g., one on a page that is no longer live). Issues like either of these are common in advertising data and could massively influence our average estimates; further, these may or may not be observable to the econometrician.

Our response to this is threefold. First, to make our estimates more robust to outliers, our main results all quantile based. Second, rather than making ad hoc assumptions about what experiments should and should not be included in the analysis, we instead cast a wide net; rerunning our analyses dropping different extreme percentiles yields quite stable estimates. Finally, and relatedly, we aimed to make our results as representative as possible for the population of advertisers who are using offsite data properly, and hence erred on the side of inclusion rather than exclusion.

6.2 Methodology

Conditional on our observed distributions of treatment effects and standard errors, we want to estimate the true, underlying distribution of effects. Further, we want to do so in as flexible a way as possible, especially given the nonstandard distributions we are dealing with. We start by articulating the general problem at hand and then show how recently developed deconvolution methods can allow us to flexibly estimate our quantities of interest under minimal assumptions.

Suppose from a set of N experiments we observe a set of treatment effects X_1, X_2, ..., X_N. Each X_i is a noisy measure of experiment i's true, unobserved treatment effect Θ_i, and we assume that the Θ_i are distributed according to an unobserved distribution $g(\theta)$. We are interested in making inferences on g based on our realized X_i's.

Assume that the unobserved distribution of treatment effects are drawn iid from g ($\Theta_i \overset{\text{ind}}{\sim} g(\theta)$ for all $i \in \{1, 2, ..., N\}$). Further, assume that each X_i is drawn independently from Θ_i according to a known distribution p_i: $X_i \overset{\text{ind}}{\sim} p_i(X_i \,|\, \Theta_i)$. Note this specifies a hierarchical distribution: first the Θ_i are drawn from g, and then conditional on the Θ_i's, we draw our realized X_i's.

Our approach here is that of Efron (2014, 2016), whose insight is given this set up, if we assume a broad exponential family of models for g, not only can we attain flexible functional forms, but the optimization problem also becomes tractable. We now lay that approach out, introducing further notation.

Specifically, assume the support of g is a finite discrete set $\mathbb{T} = \{\theta_1, ..., \theta_m\}$. (This assumption is not strictly necessary, but it eases the analysis.) This makes the prior $g(\theta)$ an m-vector $g = (g_1, ..., g_m)$ that specifies the probability g_j on θ_j. We assume:

$$g(\alpha) = exp\{Q\alpha - \phi(\alpha)\} \tag{1}$$

where α is a p-dimensional parameter vector and Q is a known $m \times p$ structure matrix. Denoting by Q_j^T the jth row of Q, we have that the jth component of $g(\alpha)$ is

$$g_j(\alpha) = exp\{Q_j^T\alpha - \phi(\alpha)\} \quad \text{for } j = 1,2,...,m \tag{2}$$

where $\phi(\alpha)$ normalizes $g(\alpha)$ to make it a probability distribution:

$$\phi(\alpha) = \log \sum_{j=1}^{m} exp(Q_j^T\alpha) \tag{3}$$

In our estimation, we will follow Narasimhan and Efron (2020) in letting Q be basis matrix for natural cubic splines over $\mathbb{T}$ with degrees of freedom p. We let interior knots be distributed evenly across percentiles of the treatment effects and explore model fit via BIC over a wide range of p values (more details later).

Let

$$p_{ij} = p_i(X_i \,|\, \Theta_i = \theta_j) \tag{4}$$

denote the probability that X_i is realized if $\Theta_i = \theta_j$, and let P_i be the m-vector of probabilities for X_i across all possible values of θ_j: $P_i = (p_{i1}, ..., p_{im})^T$.

Importantly, note the i subscripts on $p_i(X_i \,|\, \Theta_i)$; this means that each experiment can have a different conditional probability distribution over the observed treatment effects. In our setting this comes up since we not only observe X_i but we also observe an estimate of the variance, $\hat{\sigma}_i{}^2$, that varies by i. In our set up, we assume $p_i(X_i \,|\, \Theta_i) \sim N(\Theta_i \,|\, \hat{\sigma}_i{}^2)$, or that the variances are known (as is standard in meta-analyses, e.g., DellaVigna and Linos (2022)). Note this is not assuming normality of g, but rather normality of the distributions that generate X_i conditional on Θ_i.

Given all this set up, the marginal probability for X_i becomes:

$$f_i(\alpha) = \sum_{j=1}^{m} p_{ij}g_j(\alpha) = P_i'g(\alpha) \tag{5}$$

and hence the log likelihood function is $l_i(\alpha) = \log\, P_i'g(\alpha)$. We can use maximum likelihood to generate an estimate $\hat{\alpha}$ for α.[15]

[15] Note this approach is sometimes called g-modelling to emphasize that we are interested in the shape of the prior; most empirical Bayes methods are f-modelling in that they are interested in the marginals.

6.2.1 Implementation

We will explain the implementation for the business as usual case, as the details are the same for both meta-analyses.

First, we defined a discrete state space $\mathbb{T}$ for the support of $g(\theta)$. Given the wide range of treatment effect estimates, discretizing the entire range from the smallest to largest treatment effect estimate is computationally challenging. Hence, we defined a grid of bin size 0.001 over the range from the smallest to the largest treatment effect estimates of between the 1st and 99th percentiles, and then added singleton points for the treatment effect estimates outside of that.[16]

As aforementioned, Q was chosen to be to the basis matrix for natural cubic splines with degrees of freedom equal to p. To minimize the risk of rounding errors we standardized the basis matrix to have columns mean zero and sum of squares one. Intuitively, we are defining a flexible basis for splines across the parameter range and then using maximum likelihood to estimate the weights across the bases. As we vary p we change the number of knots, thereby adding increased flexibility at the risk of overfitting.

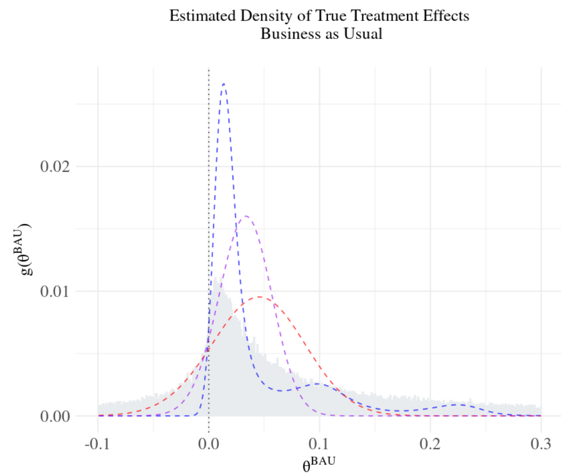
Note in our set up the only free parameter is p - we ran models from a range of starting values over $p \in \{10, 20, 30, ..., 200\}$ and for each run computed the Bayesian Information Criterion (BIC) (Schwarz 1978). At higher levels of p the estimates become quite stable, though for each we chose the model with the lowest BIC for the results (p=160, 130 for the business as usual and within product difference, respectively). Note these correspond to p-1 knots spaced evenly across percentiles of our treatment effects.

6.3 Results

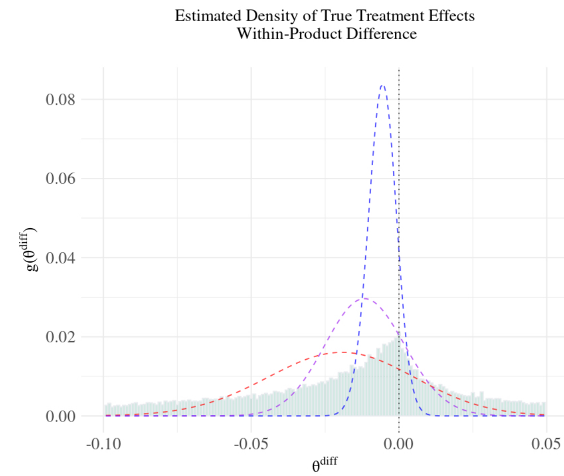
6.3.1 Overall

Given this implementation, below we plot the distribution of treatment effects again, with our fitted models overlaid. We also include models with lower BIC's and degrees of freedom to capture how model fit improves as we add more flexibility.

[16]Rerunning the analyses restricting just to the range between the 1st and 99th percentile treatment effects, the median point estimates were virtually unchanged, though the mean was reduced, as one would expect given the direction of the skew in the underlying distribution. Similarly, running with smaller bins sizes did not change the results meaningfully.



(a) Business as usual



(b) Within-product difference.

Figure 7: Distribution of observed treatment effects for each of advertising effectiveness today and the within-product difference (bin width of 0.001).

For more detail, below we also report out summary statistics of our selected specifications:

Table 5: Summary statistics of estimated distributions of true treatment effects (g). To aid exposition, we also add superscripts to the respective distributions and display the inverse of our point estimates to recover the number of incremental customers per dollar under business as usual.

	10th	25th	50th	75th	90th	Mean
g^{BAU}	0.0058	0.0118	0.0228	0.0788	0.1988	0.1428
In dollars:	(\$172.77)	(\$84.83)	(\$43.88)	(\$12.69)	(\$5.03)	(\$7.00)
g^{diff}	-0.0122	-0.0092	-0.0062	-0.0022	-0.0002	-0.0057

A few comments on these results. First of all, note in both meta-analyses how lower-dimensional exponential distributions (of which the normal is an example) fit the data quite poorly and using those could lead to substantial bias. This speaks to the value of our flexible approach in estimating the underlying distributions.

Second, while estimating the returns to advertising has historically produced a large number of null results, given our sample size and approach we can speak to the distribution of effects across advertisers. We find 97.3% of distribution of estimated true effects lies above zero (($g^{\mathrm{BAU}} > 0$); similarly, 90% of the distribution of estimated effects of signal loss lie below 0. Importantly, however, these do not address whether advertising is profitable for these businesses or whether that loss of effectiveness is economically meaningful for the firms.

The median business as usual result evaluated at the median difference result is an increase from $43.88 to $60.19, which is about a 37% increase. This is a substantial change, one which might have effects on both the intensive and extensive margins of advertising for firms. Back of the envelope, the median total weekly ad spend in our sample was $1,259, which at our median estimate of ad effectiveness is about 29 new customers per week; holding spend constant, the median advertiser after loss of offsite data would then only generate 21 customers per week.

Finally, in thinking about the generalizability of these results, we note compared to other studies a primary benefit of our sample is its representativeness. Since only a small minority of advertisers use measurement tools, past work generally relies either on quasi-experimental evidence or experimental evidence from a small subset of advertisers (e.g., Gordon et al. (2019)). Here, in contrast, there is minimal selection bias, no publication bias, and we can feel quite confident that these results are representative to a high degree of the population of advertisers who are using offsite data in this central way.

6.3.2 Heterogeneity

Conditional on our sample, we can also explore how effects differ across demographics. To that end, in this section we explore how the effects differ by advertiser size and vertical. Specifically, we rerun our main analysis restricting to each group so we can see, for example, how effective advertising is today for Retail spenders as well as how much that would change under loss of offsite data.

We start by rerunning our main analysis for businesses that are characterized as "small" and "large" internally. Again, this is largely determined by historical ad spend and is correlated with offline firm size, albeit imperfectly. Here, we reran our core specification over $p \in \{50, 100, 150, 200\}$ for each of business as usual and the within product difference; we again selected the best model by the BIC for each.

First, to visualize what this looks like, below we plot the fitted distributions for each of small and large businesses (Figure 8). Here, the blue distributions are those for small businesses and the dashed line denotes business as usual (hence, red is for large businesses and dotted is the within-product difference). We can see here that the distributions for small businesses are shifted out in each case: their ads tend to both generate more incremental customers per dollar and will be hurt more by the data loss.

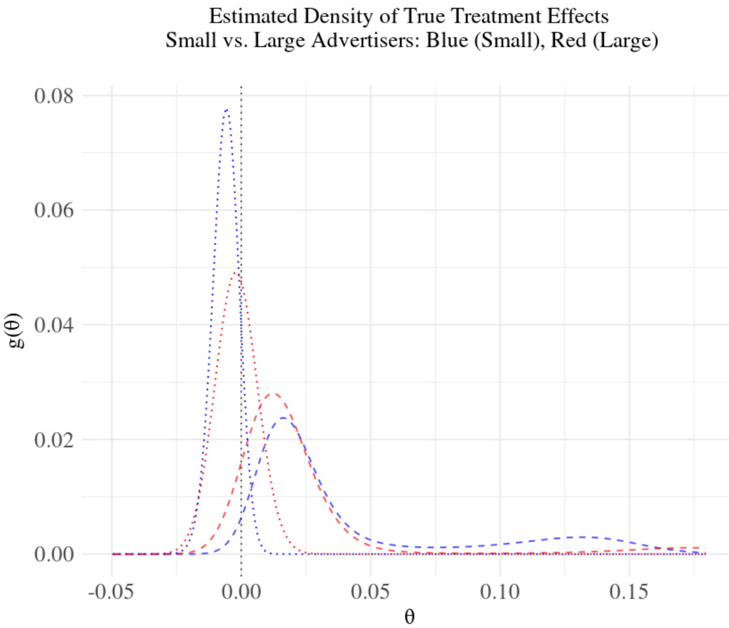


Figure 8: g distributions for internally classified "small" and "large" businesses, business as usual and within-product change.

For more detail, below we report out summary statistics of the resulting estimated g distributions.

Table 6: Summary statistics for g distributions for small and large advertisers, number of incremental converters per dollar and effects of signal loss.

	10th	25th	50th	75th	90th	Mean
Small Advertisers						
g^{BAU}	0.0064	0.0134	0.0249	0.0914	0.1574	0.1326
In dollars:	(\$156.41)	(\$74.66)	(\$40.17)	(\$10.94)	(\$6.35)	(\$7.54)
g^{diff}	-0.0138	-0.0098	-0.0068	-0.0028	0.0002	-0.0070
Large Advertisers						
g^{BAU}	-0.0005	0.0065	0.0155	0.0275	0.1295	0.1571
In dollars:	(-\$1,881.05)	(\$154.6)	(\$64.65)	(\$36.41)	(\$7.72)	(\$6.37)
g^{diff}	-0.0125	-0.0075	-0.0025	0.0035	0.0085	-0.0019

In the table we can see numerically that across all the percentiles above, small advertisers both get more customers per dollar spent and are more hurt by the loss of offsite data. This is even despite the fact that many small advertisers are newer to the platform, often generate lower quality ads, and are less likely to use the platform to its fullest capacity. This result is consistent with diminishing returns of advertising.[17] Relatedly, we note that as the amount of historical ad spend increases, advertisers are more likely to shift from direct response to brand advertising. The businesses who allocate more of their budget to driving purchases are disproportionately the smaller spenders, which is consistent with a model where costs of driving such lower funnel outcomes increases and it becomes relatively more attractive to target more upper funnel actions (e.g., brand sentiment, unaided awareness).

Looking at the magnitudes for small versus large businesses, we see that the median cost per incremental customer for small businesses evaluated at the median difference would rise from \$40.17 to \$55.23, a 37% increase; for large businesses the corresponding amount is \$64.65 to \$77.01, a 19% increase. Percent-wise, this indicates small advertisers would be hit nearly twice as hard as large ones (though the effects are closer in absolute terms). Note also that the smaller advertisers have a long right tail whereas the distribution for the larger advertisers is more compressed with a smaller variance.

Next, we look at results for our three largest verticals: CPG, E-commerce, and Retail:

Table 7: Summary statistics of estimated g distributions by vertical.

	10th	25th	50th	75th	90th	Mean
Consumer Packaged Goods						
g^{BAU}	0.0047	0.0107	0.0187	0.0357	0.1217	0.0559
In dollars:	(\$212.88)	(\$93.48)	(\$53.48)	(\$28.01)	(\$8.22)	(\$17.88)
g^{diff}	-0.0133	-0.0103	-0.0073	-0.0043	-0.0013	-0.0071
E-commerce						
g^{BAU}	0.0043	0.0143	0.0273	0.0453	0.2013	0.0818
In dollars:	(\$231.92)	(\$69.87)	(\$36.61)	(\$22.07)	(\$4.97)	(\$12.23)
g^{diff}	-0.016	-0.012	-0.0085	-0.004	-0.001	-0.0079
Retail						
g^{BAU}	0.0052	0.0192	0.0362	0.1082	0.2282	0.1909
In dollars:	(\$193.81)	(\$52.19)	(\$27.66)	(\$9.25)	(\$4.38)	(\$5.24)
g^{diff}	-0.0237	-0.0177	-0.0117	-0.0057	-0.0007	-0.0116

[17]We note in practice there may be multiple reasons for this: for example, it could be that the Offsite Conversion Optimization finds people with the highest probability of converting conditional on exposure, and the shape of that conditional distribution generates the diminishing returns; alternatively, the conditional probability of conversion could be constant for many users but there could be increasing search costs to finding them. From the advertiser's perspective though both would have the same effect of diminishing returns from their ad spend.

Again computing the increase in costs from the median estimates, we find substantial increases in costs above and beyond that we observed for our main effect for these verticals: 64%, 45%, and 48% for CPG, E-commerce, and Retail, respectively. These vertical represent some of the heaviest users of not only digital advertising but also offsite data, consistent with these large effect sizes. We note the result is particularly concerning for E-commerce firms, many of whom lack a physical storefront and rely on online ads to bring consumers to their business.

6.3.3 Long Term Effects

Thus far we have restricted attention to incremental customers over the course of our experiment. However, not all customers are equal for advertisers: for example, it could be that optimization with offsite data is better at finding impulsive buyers or consumers who are likely to be tricked into purchasing. In contrast, a consumer who selects into being a long term buyer is higher value for the advertiser. Further, if an ad induces a consumer to become a long term purchaser, that is consistent with a larger benefit on the consumer side than if the ad were to induce either zero or one purchases.

To study this question further, we looked at purchasing behavior across our treatment groups six months after our initial experiment ran. Given the initial random assignment, we can identify the effects on long term purchasing behavior, and see how that varies across campaigns with or without offsite data.

We again reran our models over a wide range of p values and selected via BIC (p=70,220 for BAU and the within-product difference, respectively). Summary statistics of these distributions are provided below.

Table 8: Summary statistics of estimated distributions of true treatment effects (g) for long-term effects. Specifically, the cost per incremental customer in a week-long period six months after our initial experiment.

	10th	25th	50th	75th	90th	Mean
g^{BAU}	0.0010	0.0030	0.0055	0.0080	0.0105	0.0057
In dollars:	(\$1,033.59)	(\$336.98)	(\$182.90)	(\$125.51)	(\$95.53)	(\$174.28)
g^{diff}	-0.0029	-0.0024	-0.0014	-0.0006	0.0001	-0.0013

In the purchase-optimized group, the median cost per incremental customer six months in the future is \$182.90. Again looking at the within-product difference distribution and repeating our usual calculation there, we get that under click-optimized that would increase to \$245.35, which is roughly a 34% increase. We note this is a similar magnitude to the short run estimate. In other words, per dollar spent, purchase-optimized generates substantially more longer-term customers (under this definition) than click-optimized.

7 Conclusion

The focus of our paper is on estimating the value of offsite data to advertisers using a large sample of advertisers on Meta. Offsite data is both believed to be amongst the more important in digital advertising and policy relevant as ongoing regulatory and product changes threaten the ability of advertisers to use this data. To estimate first order, partial equilibrium effects of this data on ad effectiveness, we conducted a large-scale experiment with more than 100,000 advertisers.

We find ad effectiveness would be substantially hampered by loss of this data. The median product evaluated at the median loss would experience as 37% increase in costs per incremental customer. There is further evidence smaller scale advertisers will be disproportionately hurt, as would advertisers in CPG, E-commerce, and Retail.

The debate over inclusion or exclusion of data in ad delivery is often framed as a binary choice between privacy for consumers and ad effectiveness for firms. While here we have provided evidence on the magnitude of the latter, we note that such a binary framing is misleading. New advances in privacy-enhancing technologies, differential privacy, and federated learning may allow both objectives to at least be partially met. However, such technologies are still developing, and before they are ready, policymakers and companies must weigh the tradeoffs in altering the offsite data ecosystem.

References

Ethan Cramer-Flood. Worldwide digital ad spending 2021. Technical report, eMarketer, 2021. URL https://content-na1.emarketer.com/worldwide-digital-ad-spending-2021.

Robert Hackett and Declan Harty. Apple's ad changes wiped $142 billion off snap, facebook, and other online ad giants. Technical report, Forbes, 2021. URL https://fortune.com/2021/10/22/apple-snap-facebook-earnings-google-twitter-pinterest-ad-tracking/.

Christine Moorman. Cmo survey. Technical report, Duke University, February 2022. URL https://cmosurvey.org/wp-content/uploads/2022/02/The_CMO_Survey-Highlights_and_Insights_Report-February_2022.pdf.

Anna Peck. 2022 small business advertising report. Technical report, Visual Objects, March 2022. URL https://visualobjects.com/advertising/blog/small-business-advertising-2022.

Werner Geyser. The state of the direct-to-consumer (dtc) industry – 2022. Technical report, Influencer Marketing Hub, March 2020. URL https://influencermarketinghub.com/direct-to-consumer-industry/.

Bradley Efron. Empirical bayes deconvolution estimates. *Biometrika*, 103(1):1–20, 2016.

Bradley T Shapiro, Günter J Hitsch, and Anna E Tuchman. Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica*, 89(4):1855–1879, 2021.

Avi Goldfarb and Catherine Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 2011a.

Randall A. Lewis and Justin M. Rao. The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics*, pages 1941–1973, 2015.

Garrett A Johnson, Randall A Lewis, and Elmar I Nubbemeyer. Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6):867–884, 2017.

Brett R Gordon, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. *Marketing Science*, 38(2):193–225, 2019.

Alexander Bleier and Maik Eisenbeiss. Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, 34(5):669–688, 2015.

Navdeep S Sahni, Sridhar Narayanan, and Kirthi Kalyanam. An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research*, 56(3):401–418, 2019.

Omid Rafieian and Hema Yoganarasimhan. Targeting and privacy in mobile advertising. *Marketing Science*, 40(2):193–218, 2021.

Garrett Johnson, Randall A Lewis, and Elmar Nubbemeyer. The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. *Available at SSRN 2701578*, 2022.

Garrett Johnson. The impact of privacy policy on the auction market for online display advertising. 2013.

Jian Jia, Ginger Zhe Jin, and Liad Wagman. The short-run effects of gdpr on technology venture investment. Technical report, National Bureau of Economic Research, 2018.

Avi Goldfarb and Catherine E Tucker. Privacy regulation and online advertising. *Management science*, 57(1):57–71, 2011b.

Samuel Goldberg, Garrett Johnson, and Scott Shriver. Regulating privacy online: An economic evaluation of the GDPR. *Available at SSRN*, 2021.

Stefano DellaVigna and Elizabeth Linos. Rcts to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90(1):81–116, 2022.

Bradley Efron. *The Bayes deconvolution problem.* Division of Biostatistics, Stanford University Stanford, 2014.

Balasubramanian Narasimhan and Bradley Efron. deconvolver: A g-modeling program for deconvolution and empirical bayes estimation. *Journal of Statistical Software*, 94:1–20, 2020.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, pages 461–464, 1978.