Learning, Sophistication, and the Returns to Advertising: Implications for Differences in Firm Performance*

Steven Tadelis† Christopher Hooton‡ Utsav Manjeer‡ Daniel Deisenroth‡

Nils Wernerfelt‡ Nick Dadson§ and Lindsay Greenbaum§

April 2023

Abstract

Why do establishments exhibit wide variation in their productivity and profitability? Can variation in returns to advertising help answer this question? We present results from a large field experiment on Facebook and Instagram that documents variance in advertisers' ability to generate returns to advertising. We focus on campaigns aimed at boosting sales and tie advertising expenses to revenues for each advertiser. We find that spending on advertising led to significant increases in revenues, number of purchases, number of purchasers, and number of conversions. The heterogeneity in these results by expenditure, age, and engagement documents patterns consistent with learning by doing and variance in how sophisticated advertisers are. Advertisers who engage in more learning activities and more sophisticated data collection exhibit the highest returns and are more likely to continue their activities over time, suggesting that differences in advertising effectiveness may account for some of the variance in productivity across firms.

---

* We thank Reed Walker for fruitful discussions and are grateful to Matt Backus, Mitch Hoffman, and Chad Syverson for detailed comments and suggestions. We also thank Hallie Feibel and Aditya Mohan for their research assistance. The authors are listed in order of contribution, with C.H. and U.M contributing equally. C.H. and U.M. led the project internally at Meta Inc., and implemented the experiment; C.H., U.M., and N.D. analyzed and validated the data. D.D. and N.W. were the joint internal principal investigators. S.T., L.G., and N.D. were the joint external principal investigators, and were compensated for parts of their work under a contract between Meta Inc. and Analysis Group, Inc. All authors contributed to the development of the paper itself.

† University of California Berkeley Haas School of Business, NBER and CEPR (corresponding author).

‡ Meta Platforms, Inc.

§ Analysis Group, Inc. The views presented in this article are those of the authors and do not represent any opinions or positions of Analysis Group or its affiliated companies.

1 Introduction

Large and persistent differences in productivity levels across businesses are ubiquitous, a fact that has shaped research in several fields such as industrial organization, macroeconomics, and international trade to name a few. The most productive establishments can be twice to five times more productive than the least productive ones in the same single industry (e.g., Syverson, 2004; Hsieh and Klenow, 2009), and these differences are persistent over time (e.g., Foster, Haltiwanger, and Syverson, 2008).

These observations lend support to the theoretical view that some firms are inherently more productive than others and productivity differences are a consequence of this productivity variance (e.g., Jovanovic, 1982; Hopenhayn, 1992; and Ericson and Pakes, 1995.) As a consequence, when market conditions change, such as demand shocks or entry by new firms, the least efficient firms will be forced to exit, while the more efficient ones will survive the storm. As Syverson puts it, "some producers seem to have figured out their business (or at least are on their way), while others are woefully lacking." (Syverson, 2011 p. 327)

The literature has focused primarily on production-process-related methods as the source of heterogeneity across firms. Beyond production-processes, productivity differences across establishments may be caused by differences in their human capital (e.g., Abowd et al. 2005), in human resource practices and incentive structures (e.g., Lazear, 2000; Ichniowski, Shaw, and Prennushi, 1997), as well as in management practices (e.g., Bloom and van Reenen, 2007). Production alone, however, which is admittedly the core focus of textbook economic models, is only part of what makes a business successful; without marketing and sales activities, most firms would have trouble selling their goods and services. As Bronnenberg, Dubé, and Syverson (2022) observe, investments in brand activity, such as marketing, advertising, and sales, are seldom if ever considered as drivers of variation in performance as these are typically not recorded in conventional data sources, and often simply hard to measure.

In this paper, we take a first step in filling this gap by focusing on online advertising activities of over 200 thousand establishments in the United States, and documenting how varied the returns to advertising are. We begin with a large-scale field experiment conducted at Meta, Inc., in which we utilize tools that allow us to measure revenues on the advertisers' webpages. We randomly hold out a small fraction of targeted Facebook and Instagram users from seeing ads to measure the causal returns to advertising. We then measure how returns to advertising vary across advertisers across two dimensions of advertiser characteristics: first, with respect to how advertisers interact with the advertising platform, which we interpret as a form of learning by doing (Arrow, 1962), and second, with respect to the quantity and quality of data that they use, which we interpret as a form of sophistication. We then explore how these differences impact activities of these businesses on the advertising platform, which we posit to be related to their long-term survival.

We first find that ad spending on these platforms has a significant positive effect on advertiser revenues, as well as their number of purchases, purchasers, and conversions, which establishes that advertising does impact revenues in a meaningful way. Next, we stratify our sample by campaign and advertiser experience and find that returns are significantly higher for campaigns and advertisers with more experience. Importantly, this relationship is present only for the subset of relatively engaged advertisers that update their campaigns over time and is absent for unengaged advertisers. This is consistent with advertisers learning how to increase the effectiveness of advertising, rather than solely relying on the familiarity of, or improvements in, Meta algorithms. Furthermore, we find that engaged advertisers who use more data-generating tools see higher returns to engagement, consistent with learning activities and richer information acquisition being complementary. Last, we show that advertisers who either exhibit more learning or more sophistication are significantly more likely to actively advertise months after the experiment, with those more engaged and more sophisticated exhibiting the highest likelihood of survival.

Our experiment was intentionally designed to measure returns to advertising in dollar values, consistent with the revenue-based productivity measure (TFPR) that is standard in the literature (e.g., Foster, Haltiwanger, and Syverson, 2008). To do this, the experiment consisted of all ad campaigns on Facebook and Instagram that were optimized to boost purchases[1], were run by advertisers who had spent money on ads in the 90 days preceding the experiment, and were tied to a "Meta Pixel", which is the technology that allowed us to record revenues on the advertisers'

---

[1] As we describe in more detail later, we restrict to campaigns where the advertiser specified sales as an objective. This is an added strength of our data: in many advertising studies, it is hard for the researcher to know the advertiser's objective (e.g., brand awareness, registration, or some other direct response), which complicates efforts to measure effectiveness. Here, we know what the goal is and can directly measure against it.

websites. Section 2 offers a detailed description of how advertising on the Meta platforms work and Section 3 describes the fine details of the experimental design and the data it generated.

In total, our data contains information from approximately 3.94 billion user-ad opportunity pairs from over 700,000 ad campaigns run by over 200,000 advertisers in 25 industries (in the U.S.); these campaigns cover a large proportion of annual U.S. ad spending on Meta.[2] The size of our experiment allows us to aggregate our data to the level of each ad campaign-experimental group combination, which reduces the variance in the outcome variables relative to their individual-level counterparts while still giving us sufficient data to estimate the returns to ad spend. Further, our experimental universe was purposefully broad, containing ad campaigns of a variety of different sizes from a variety of different industries. This assuages some potential selection concerns.

Not only do we find large and persistent differences across firms in advertising, but there are also predictable differences within firms, namely, there seems to be substantial learning by doing, as well as significant returns to being more sophisticated data users. We demonstrate this by first creating benchmark results on the average returns to advertising. We find that each dollar spent on ads yields $3.31 in revenues, and that ad spending drives a 25 percent increase in the number of purchases, a 13 percent increase in the number of unique customers/purchasers, and a 73 percent increase in conversion events. We then use the return on ad spend as our primary benchmark of interest to further examine performance variation and analyze these results across different advertiser groups. We find a distinct relationship between advertiser engagement and higher returns, particularly stemming from the usage of more data features. This is shown across a variety of advertiser characteristics—including historical ad spend, the number of ads and campaigns previously run, and the age of advertisers—where we see improvements in revenue performance of 22-122% for advertisers with above median levels of each as compared to those below the median. This latter finding suggests complementarities between learning activities and data inputs that support learning.

We believe that our results shed light on the question of why some establishments exhibit stronger performance than others. Taking our benchmark results as evidence that advertising activities help drive a company's revenue, the heterogeneous patterns in our data show that there is lots of

[2] While we cannot provide exact figures due to privacy requirements, it is sufficiently high for external validity.

variance in how well companies use advertising. Because one dimension of our analysis focuses on learning as a source of variation, our paper contributes to the empirical literature on learning by doing that, by and large, has focused on increased productivity in industries with relatively constant technologies (e.g., Benkard, 2000; Hendel and Spiegel, 2014; and Levitt, List, and Syverson, 2013). After all, the observation in Arrow (1962) that learning by experience in production is central to economic growth was inspired by Lundberg (1961), who documented a gradual increase in a steel-mill's output without any capital investments. These studies focus on a single firm to precisely measure output and inputs with precision. In our setting, the learning is not about the production of goods or services, but instead about using a key ingredient in generating sales regardless of the product or service: advertising to potential consumers. Therefore, we can aggregate performance from many advertisers across many industries, as our measure of input and output are in U.S. dollars, rather than a particular product. In addition, recent work has documented patterns of learning by firms that are not about production itself. Doraszelski et al. (2018) document patterns of learning in bidding behavior by firms in a newly opened electricity auction market, and Backus et al. (2023) document patterns of learning how to better use language in an online bargaining setting.

We also contribute to the recent literature on measuring the returns to digital advertising. Identifying the returns to advertising is challenging for several reasons. First, most online advertisers use observational, non-experimental methods to determine the effectiveness of their advertisements, a practice plagued by endogeneity issues, and that can only be corrected for using well designed experiments (e.g., Lewis, Rao, and Reiley, 2011; Blake, Nosko and Tadelis 2015; Gordon et al. 2019).[3] Second, even using experimental methods to identify the returns to advertising is often challenging. For example, economically significant effects of advertising are often statistically insignificant, even in large field experiments with rich covariates, because of the large variance in user-level sales. The sample sizes and implied expenditure required to reliably detect the effectiveness of ad campaigns can easily exceed the amounts feasible for many advertisers (Lewis and Reiley 2014; Lewis and Rao 2015). Interestingly, Lewis and Reiley (2014) study a large retailer and demonstrate that 93% of the increase in purchases due to advertising occurred in brick-and-mortar stores. Studies that exclusively examine online sales, like ours, may

[3] See Johnson (2022) for an extensive overview of field experiments in online display advertising.

therefore underestimate the effect of online advertising, meaning that our estimates are a lower bound on how effective advertising is in our setting.

Digital advertising was only introduced in the early 2000's, with tools and interfaces that had never existed before for business owners and marketing leaders. In the past decade, digital advertising spending in the U.S. alone has grown more than five-fold from $36.6 billion in 2012 to $189.3 billion in 2021.[4] The upshot of our analyses is that it is incumbent upon establishments who rely more and more on digital advertising to engage in experimentation and learning to get the best bang for their buck. What's more, unlike other non-technological aspects of productivity, such as labor force composition, compensation incentives, and management practices, it is much easier to change aspects of advertising activities and it is much faster to measure the impact of these changes. The challenge remains for establishments to have the right talent and an attitude of sophisticated learning, which may be difficult in a highly competitive market for scarce talent.

2 Advertising on Facebook and Instagram

2.1 Advertising Campaigns and Objectives

Advertisers must first create a Facebook Business page or an Instagram account to advertise on Facebook and Instagram. This process allows advertisers to use the Meta Ads Manager tool, through which they can create and manage advertising campaigns.

An advertising campaign is a collection of related ads served concurrently by an advertiser. Although the format of ads may vary within a campaign, their overall message is generally consistent. Importantly, when creating an ad campaign, the advertiser must first specify an objective—such as improving brand awareness, increasing visits to the advertiser's business page,

[4] See "Internet Advertising Revenue Report." April 2022. https://www.iab.com/wp-content/uploads/2022/04/IAB_Internet_Advertising_Revenue_Report_Full_Year_2021.pdf

or boosting sales—that the campaign is designed to meet.[5] Once an ad campaign is active, Facebook and Instagram display ads as users scroll through their feeds.[6]

As mentioned earlier, ad campaigns can be optimized to achieve specific objectives. Our focus in this paper is exclusively on ad campaigns that were designed by the advertiser to increase sales. However, irrespective of the campaign's objective, there are several steps between the advertiser registering a campaign on the Ads Manager tool and the campaign actively delivering ads to users. While each element plays an important part in the delivery of ads to platform users, advertisers can actively manage features in most of these steps and thereby influence the ultimate delivery of the campaign. Subsections 2.1.1-2.1.3 provide some context on these different aspects of advertising on Facebook and Instagram that will prove useful in understanding and interpreting our key results.

2.1.1 Creative Design

After an advertiser has defined their campaign objective (e.g., increasing website visitors or increasing sales), the next steps involve designing the specific advertising creatives to be used in achieving the campaign's goals. Ad *creatives* are the actual media content delivered to users, which advertisers can design creatives from scratch or, alternatively, they may review and customize simple drafts generated by Meta's automated creative tools. These creatives can range in complexity from simple text to professionally-produced video to interactive digital collections, and undergo review by Meta to ensure they comply with policy guidelines.[7]

Previous research has illustrated that better design and aesthetic features of online content leads to greater user engagement and sales (Li and Xie, 2020; Zhang et al., 2022). Advertising campaigns may, therefore, perform better *a priori* when advertisers invest more time and resources into the

---

[5] These objectives are closely tied to the concept of the "Digital Marketing Funnel" that captures the typical buying journey of consumers starting from knowing a company's brand, through visiting their site, and finally making a purchase. The funnel concept seems to have been first introduced by Strong (1925), who credits the idea to Elias St. Elmo Lewis, creator of the "AIDA model" (Attention, Interest, Desire Action.) See, e.g., Johnson, Lewis and Nubbemeyer (2015) for a review of the concept in digital advertising.

[6] Ads appear in other locations as well – for example, on the right hand side of a user's Facebook feed while on desktop browsers – but in-feed ads are a canonical example for the reader to bear in mind.

[7] There are five supported Meta ad media formats: image, video, carousel (a collection of images and or videos), instant experience, and collection. For more, see the article "Types of Meta ad formats" in the Meta Business Help Center (Meta, n.d.).

aesthetic development of the ad creatives. Advertisers may observe that an ad with a blurry image, for example, achieves fewer conversions, and devote effort to develop a clearer, high resolution image for the ad campaign with the aim of increasing conversions.

2.1.2 Targeting

In addition to developing high quality imagery for their ads, advertisers may choose specific audiences to target. Advertisers can target specific audiences based on user characteristics and actions while creating ad campaigns on Facebook and Instagram. This reflects one of the major advantages of digital advertising over traditional advertising on mass media—digital ads allow advertising businesses to target relevant consumers, and reach "niche" markets at relatively low costs. For example, an advertiser could target a user audience consisting of females between the ages 30–40 who have clicked on sneaker ads on Facebook. Meta further allows advertisers to build "custom audiences", using information that advertisers may collect on their customers, and "lookalike audiences", comprising on-platform users who share the demographics, behaviors, and interests of the advertiser's custom audience.[8] For example, the advertiser's target user audience described above may consist only of members of the advertiser's email list and other users who share similar characteristics.

Advertisers may also, at any point, choose to opt-in to Meta's targeting expansion feature, which is an additional feature beyond the use of target audiences. The targeting expansion feature uses algorithms to expand the campaign's reach beyond the pre-specified targeting parameters with the goal of reaching a larger audience of users that are likely to interact with the ad. It does so, however, not by creating lookalike audiences, but rather by adjusting the existing targeting criteria other than location, age, gender, and pre-specified exclusions, if any.

2.1.3 Auctions, Bidding, and Budgets

Once advertisers have chosen their campaign objectives, designed ad creatives, and specified target audiences, they must decide on their campaign budget and ad auction bidding strategies. These strategies, taken together with other campaign parameters, are instrumental in delivering the ads to users most likely to engage with those ads. As with most online advertising, Meta's ad

[8] See the articles "About lookalike audiences", "About Custom Audiences", and "Audience ad targeting" at the Meta Business Help Center (Meta, n.d) for more details.

algorithm uses an ad auction to determine which ad to deliver to a given user at a given point in time. However, before launching the ad campaign to deliver ads to users, the advertiser must determine how much to spend on the ad campaign, as well as how to distribute the budget across time and ad auctions.

Advertisers can flexibly choose across several options pertaining to each of these parameters. For example, advertisers can set an overarching budget for the overall campaign or budgets for individual ads, where the former allows for greater automation in optimization and allocation of ad budgets across ad auctions, the latter gives advertisers greater control in ad delivery.[9] Advertisers can also choose between daily budgets—which allows advertisers to spend a similar amount each day throughout the duration of the campaign—and lifetime budgets—which set a hard cap on the overall campaign budget, but may vary at the daily level with a soft cap to allow for optimal placement. Further, advertisers can choose between different bidding strategies for the ad auction, which we elucidate below in greater detail. In particular, advertisers decide across a spend-based bidding strategy to maximize an outcome value per bid, or a goal-based bidding strategy to maintain a cost or value-based goal throughout the campaign, or manually set maximum bid values for ad auctions. Finally, at any point in time, advertisers can change their campaign budgets and bid strategies. Meta's ad algorithm takes account of the advertiser's specified budget and bid strategies, and enters their ads into ad auctions, which are used to determine the best ad to send to a given user at a given point in time.

For each potential ad impression on Facebook and Instagram, Meta gathers the ads that include the relevant user in their audience and enters them into an ad auction. The auction is a version of a Vickrey-Clark-Groves mechanism, in which the winner only pays the minimum required to have won the auction. If, for example, advertisers decide on a campaign budget and allows Meta to bid automatically based on that budget and the advertiser's pre-specified strategies, bidding will continue for eligible ad impressions until the allocated budget has been depleted. Meta uses machine learning algorithms trained on user behavior, characteristics, and other factors to estimate the "Action Rate", which is the user's likelihood of taking the advertiser's objective-specified

[9] Technically, the budgets can be set at the campaign or "ad set" level. Ad sets are groups of similar individual ads that share settings for how, when, and where to run. For example, setting budgets at the campaign level allows Meta to automatically increase spending on those ads that are performing well with users. For more, see the article "About campaign budgets and ad set budgets (Meta, n.d.).

action. Similarly, Meta uses machine learning algorithms trained on user feedback and ad attributes to estimate the "Quality Score", which is a determination of the overall quality of the ad. The auction winner is determined based on which ad has the highest "Total Value", Meta's combination of the Advertiser Bid, the ad's Action Rate for the user, and the ad's Quality Score: Total Value = Advertiser Bid × Action Rate + Quality Score.[10] Note in this formula that users who are predicted to be more likely to undertake the advertiser-specified action (e.g., purchase the product) will have the respective bid scaled up, thereby making that ad more likely to win the auction. This is a primary channel through which ad delivery is optimized: if an advertiser wants to drive sales, endogenously their bids will be higher in auctions where the user is predicted to be likely to purchase the product, thereby helping to selectively allocate budget to likely converters.[11]

2.2 Pixels and Outcomes

As we describe earlier, ad campaigns can be optimized to achieve different objectives, such as increase visits to the advertiser's business page, boost sales, or improve brand awareness. In the following subsections, we briefly describe the outcomes advertisers are typically interested in, and how advertisers use the Meta Pixel to measure the effectiveness of their campaigns in terms of improving such outcomes.

Advertisers can obtain on-platform information on the efficacy of their ad campaigns, such as the number of ad impressions and clicks directly from the Meta Ads Manager tool. However, outcomes that occur off-platform, such as on the advertiser's own website, are measured through the Meta Pixel. Meta describes the Meta Pixel as, "a piece of code that you put on your website that allows you to measure the effectiveness of your advertising by understanding the actions people take on your website."[12] For example, a Pixel installed on an advertiser's website could record every action of a user, such as adding an item to their shopping cart or making a purchase.

[10] See Meta, 2020. https://www.facebook.com/business/news/good-questions-real-answers-how-does-facebook-use-machine-learning-to-deliver-ads.

[11] Note that there is no mention of incrementality; e.g., it's possible that optimizing delivery for purchases shows ads to users who were already going to buy the product. We take this concern seriously and, as we describe later, structure our experimental design to measure incremental effects.

[12] Even though advertisers can have multiple Pixels and campaigns, each campaign can be linked to only one Pixel. See https://www.facebook.com/business/help/742478679120153?id=1205376682832142.

There are 17 standard events that Pixels can monitor and report back to Meta.[13] Advertisers can choose to implement the monitoring and reporting of any or all of these standard events since advertisers own the web pages where Pixels are embedded. Pixels also allow advertisers two further customization options. First, advertisers can program a Pixel to monitor and report additional advertiser-customized events, such as when a user clicks on an image on the advertiser's website. Second, Pixels can be used to separately monitor "custom conversion" events, which act as filters over standard events. For instance, the Pixel can report a custom conversion event each time a user makes a purchase worth over $50.

Pixels play an important role in advertising on Facebook and Instagram. They can be used to track website visitors' actions, also known as conversion tracking. This can, in turn, be used to analyze the effectiveness of a "conversion funnel" (recall from Section 2.1) and to calculate return on ad investment. Pixels can also record multiple user actions that facilitate user targeting, enable the estimation of actions, and allow for the tracking of outcomes that can inform ad effectiveness. One of the reasons Pixels are popular amongst advertisers is that they allow advertisers to create campaigns where ad delivery is optimized for sales. Since sales usually occur off platform from Facebook and Instagram (e.g., in Safari), without Pixels, advertisers could only optimize delivery for on platform outcomes—such as clicks or video views—that may be poor proxies for ultimate sales. For advertisers interested in driving sales, Pixels are thus a logical tool to use.

As we describe in subsequent sections, Pixels also play an integral role in the execution of our experiment by recording purchases attributed to specific ads and campaigns, as well as revenues derived from those purchases.[14]

3 Experimental Design

As noted earlier, running advertising experiments can often be costly and infeasible for many advertisers. Direct response effects of advertisement treatments for individual businesses can often be too small and, like individual-level sales, too volatile (Lewis and Rao, 2015). Therefore, even

[13] The 17 standard Pixel events that can be monitored and reported back to Meta are: Add payment info, Add to cart, Add to wishlist, Complete registration, Contact, Customize product, Donate, Find location, Initiate checkout, Lead, Purchase, Schedule, Search, Start trial, Submit application, Subscribe, and View content (more details about each event on this page: https://www.facebook.com/business/help/402791146561655?id=1205376682832142).

[14] See more at: https://developers.facebook.com/docs/meta-pixel/get-started.

if firms are equipped with the considerable technical capacity required to run advertising experiments at scale, there is often insufficient statistical power to make valid, useful inferences without substantial budgetary outlays. Achieving significant statistical power requires substantially larger samples of firms and consumers, a feature fundamentally unavailable to many individual businesses.[15]

We leverage the scale of Meta's advertising platforms to design an experiment that would speak to a large and minimally selected number of ad campaigns run by advertising businesses on Meta. We use this experiment to first measure returns on advertising spend, but with the goal to exploit the scale and scope of our experiment to explore if and how advertisers learn. Our experimental universe consisted of all ad campaigns and traffic on Facebook and Instagram in the U.S. which satisfied three criteria: (i) the advertiser running the campaign had spent non-zero dollars on Meta ads in the preceding 90 days leading up to the experiment; (ii) the advertiser running the campaign had installed a Meta Pixel so that we can track activity off Meta's platforms; and (iii) the ad campaign was optimized to boost purchases, allowing us to measure revenues on the advertiser's website. These advertisers are responsible for a large proportion of U.S. ad spend on Facebook and Instagram. This allows us, as we describe in Section 3.2 below, to study over 700,000 ad campaigns run by over 200,000 advertising businesses.[16] Our experiment ran from April 11–17, 2022.

3.1 Treatment and Control Groups

Our basic experimental design leverages user-level "Ad-Eligible" and "Holdout" groups for each individual experiment. We summarize the design in Figure 1. Experiments are defined at the Pixel level: all the ads that are optimizing for purchases on a given Pixel are eligible to be included in the respective experiment.[17] We construct our experiments identically across all Pixels, greatly

[15] One consequence of this is that many industry benchmarks or heuristics are often based either on non-incremental correlations or a highly selected sample of advertisers who can afford to run experiments. Our experiment sidesteps both of these concerns.

[16] Additional details on the experimental setup, including robustness checks and the analysis of alternative budget levels, can be seen in the Appendix. Approximately one-third of our total observations, campaigns, and advertisers were used in the benchmark results while the other two-thirds were utilized in robustness checks discussed in the Appendix. See also Wernerfelt et al. (2022) who use the same experimental setup.

[17] Technically, multiple campaigns and accounts can feature ads optimized for the same Pixel, meaning that one experiment can feature ads from multiple campaigns or accounts. In these cases, all ads optimizing for a given

facilitating our analysis and avoiding experimental mismatch concerns that are common in meta-analyses. In the case of our experiment specifically, the creation of our treatment and control groups (Ad-Eligible and Holdout) used the same engineering infrastructure as Facebook's advertiser-facing lift experimentation product.[18]

To construct an experiment for a given Pixel, we take 5% of the respective campaign's target audience along with 5% of the respective budget. We further randomly subdivide this 5% of users into two groups: 90% are "Ad-Eligible users" (i.e., 4.5% of the campaign's targeted users) who are given the opportunity to see ads from the focal campaigns, and 10% are "Holdout" users (i.e., 0.5% of the campaign's targeted users), who are denied the ability to view ads from the respective campaigns. Thus, a user may be in the Ad-Eligible group for one experiment and the Holdout for another.

If an ad wins an auction and the user is in the Ad-Eligible group for that experiment, the ad is sent to the user. In this case, the minimum amount necessary to have won the auction would be deducted from the advertiser's campaign budget. If, however, the user is in the Holdout group for that experiment, the winning ad would be withheld and the second-placed ad from the auction is sent. First-placed advertisers are not charged for Holdout users' auctions, as these auctions are effectively "won" by the second-placed advertisers. The campaign budget that the first-placed advertiser would have spent on Holdout group users is allocated across users in the Ad-Eligible group. Therefore, each advertiser's overall budget is left unchanged by the introduction of the Holdout.

Compliance in this design is, by definition, perfect for Holdout users as they are never shown any ads from the campaign. However, compliance is one-sided for Ad-Eligible users. These users may not have actually viewed all auction-winning campaign ads because they may not have scrolled down far enough in their News Feed or logged off before the ad could be shown. As we describe

---

purchase event are eligible to be shown to users in the respective Ad-Eligible group and none are eligible to be shown to users in the Holdout. In our exposition we will abuse terminology slightly and refer to 'campaign' singular going forward.

[18] See Gordon et al. (2019) for another example of the use of this infrastructure.

below, our empirical specifications are based on intent-to-treat (ITT) estimates, which allows us to circumvent any selection issues affecting users' actual exposure to ads.[19]

3.2 Data and Descriptive Statistics

We compile a rich dataset using experimental data on purchases made by users from advertisers running the campaigns in our experiment, combined with on-platform data on advertisers and their advertising campaigns. The 700,000 ad campaigns from 210,000 advertisers yielded approximately 3.94 billion unique user-ad opportunity pairs and more than 8.9 million purchases across 25 industries. We aggregate that data to the level of each ad campaign and treatment for our analysis. Therefore, for example, a single observation in our analysis data contains information on the Ad-Eligible group in one advertiser's campaign for the week of the experiment. Another observation would contain information on the Holdout group for the same ad campaign.

For each campaign in the experiment, we collected data on targeted users that had the opportunity to view ads from the campaign on each day of the experiment. This included: Ad-Eligible users for which the campaign won the auction and had the opportunity to view ads; Holdout users, who would have had the opportunity to view ads; Ad-Eligible and Holdout users for which the campaign entered the auction but lost; and Ad-Eligible users for which the campaign won the auction but never saw the ads because they did not scroll down far enough or had logged off before they had a chance to see an ad.

We use information from advertisers' Pixels to measure the number of purchase events made by each user on the advertiser's website, as well as the revenues associated with those purchase events. As mentioned above, there are 17 standard events, including purchases, that Pixels can monitor and report, such as when a key page is viewed or when a registration form is completed. While not all of these events are directly related to sales, each corresponds to increased activity on an advertiser's website. With this in mind, we also gathered data on the total number of user conversion events recorded by the Pixels. And we supplement Pixel data with on-platform

[19] While a discussion of these issues is beyond the scope of this paper, interested readers should refer to Gordon et. al. (2019) for a thorough overview. Such ITT estimates are the standard ones reported out to advertisers on Meta.

information on the dollars spent on ads, as well as a host of advertiser (such as historical cumulative ad spend) and campaign-level (such as campaign launch dates) characteristics.

We aggregate the information reported by these Pixels to compile our three main outcome variables related to purchases: 1) revenue generated from purchase events, 2) the number of purchases, and 3) the number of purchasers, in each experimental group. In addition, we also compile the total number of conversions among users, which is an aggregation of conversions associated with all recorded Pixel events, and the ad budget spent by the advertising campaign, which serves as the key independent variable in our empirical specifications. Table 1 presents means and standard deviations for the dependent and independent variables used in our analyses.

Table 2 compares characteristics of advertisers in our experimental sample versus all active advertisers on Facebook and Instagram during the week of the experiment. The median lifetime number of ad campaigns, account age, and ad aesthetic scores (which represent the overall visual appeal of advertisements) for advertisers in our experiment were slightly higher. Median ad spend among advertisers in our experiment was much higher. This is consistent with research finding the cost of getting an incremental purchaser is substantially higher than the cost of other conversion outcomes, like views (Gordon, Moakler, and Zettelmeyer, 2022). Therefore, advertisers optimizing for purchasers, as in our experiment, are more likely willing to spend more on advertising. Finally, advertisers in our experiment sample had a higher number of clicks and followers across their business pages.

A final point on the data is that while there is information on Ad-Eligible groups across 714,986 campaigns, information on Holdout is available across 662,751 campaigns.[20] Some users may effectively drop out of the experiment if they are logged out during the entire experiment, become untraceable, or delete their account. In these cases, Meta algorithms are unable to assign the user an ad opportunity even though the user was assigned to an experimental group. As the Holdout group is much smaller relative to the Ad-Eligible groups by design, it is also more likely to experience a situation in which all the users in the group were unavailable for Ad Exposures, eliminating the Holdout group completely in our analysis dataset. As part of a battery of robustness

[20] Following footnote 16, we also examine effects using alternative budget allocations. When the Ad-Eligible and Holdout groups corresponding to all these allocations are taken together, our experiment consists of data from Ad-Eligible groups from 722,090 campaigns and Holdout groups from 700,441 campaigns.

checks later in the paper, we explore the sensitivity of our results to the exclusion of campaigns where this situation occurred.

4 The Overall Returns to Advertising

Our first set of analyses estimates the overall returns to advertising. In addition to serving as a benchmark result, there are multiple independent motivations for this exercise. First, overall returns is arguably the most important motivation of businesses that choose to advertise, as opposed to boosting other outcomes such as traffic and clicks. This is likely to be especially true for smaller businesses with limited resources who may not have the resources to translate other signals such as ad impressions to business success. Second, as we have noted earlier, the evidence on the returns to digital advertising is limited and, to the best of our knowledge, there are few experiments of a scale as large as ours that are equipped to accurately measure such returns.[21]

In Section 4.1 below, we first describe our empirical approach to estimate these benchmark results. We use variations of this regression specification for all our analyses throughout the rest of the paper. Then, in Section 4.2, we present results on the effect of ad spending on advertiser revenues and other related outcomes.

4.1 Empirical Strategy

Since the random assignment of users into Ad-Eligible and Holdout groups was independent of Meta's ad algorithms, the Ad-Eligible and Holdout users within our experiment were comparable in terms of demographics and the budget allocated to their auctions. Therefore, the only systematic difference between the Ad-Eligible and Holdout groups was the advertising budget. We demonstrate this statistical equivalence in user characteristics in Table 3.

This framework allows for a simple comparison across Ad-Eligible and Holdout users: any difference in observed outcomes between these groups must stem from the differences in assignment to treatment condition. We use these comparisons to examine the effect of withholding ads on outcomes.

[21] We also note there is a space that remains in the literature to examine overall advertising returns to revenue specifically on Meta.

We estimate a regression of the following form:

$$y_{gca} = \beta_0 + \beta_1\left(\text{Budget Spend}_{gca}\right) + \beta_1\left(\text{Group Size}_{gca}\right) + \theta_{ca} + \varepsilon_{gca}\,. \tag{1}$$

The dependent variable is the observed revenues, purchases, purchasers, or conversions of Ad-Eligible or Holdout group g, ad campaign c, and advertiser a during the experiment. Variable $\text{Budget Spend}_{gca}$ is the total ad campaign budget allocated to the Ad-Eligible or Holdout group during the experiment.

We include controls for the number of users in the Ad-Eligible or Holdout group, Group Size_{gca}, and campaign fixed effects, θ_{ca}. The inclusion of campaign fixed effects controls for factors across campaigns like ad destination (Facebook or Instagram or both), auction strategies, targeting, and the general quality of the ads in a campaign. In other words, our estimates exploit only within-campaign variation. We cluster standard errors at the advertiser level as outcomes may be correlated across different campaigns run by the same advertiser.

4.2 Benchmark Results

The effect of ad spending on revenues, our primary outcome of interest, is shown in Column (1) of Table 4. We treat this as our benchmark estimate for the overall returns on advertising. We find that, on average, advertisers earned $3.31 for each dollar spent on advertising.

Columns (2) - (4) of Table 4 shows estimates of the effects of ad spending on the numbers of purchases, purchasers, and conversions. We find that ad spending significantly increases the numbers of each outcome. In Column (2), we show that advertisers had 0.10 more purchases for every dollar spent, a 29% increase in purchases per dollar spent relative to a zero spend baseline. In Column (3), we see that advertisers had 0.02 more purchasers for every dollar spent, a 30% increase in purchasers per dollar spent. Compared to the effects on the number of purchases, these effects are roughly 80% smaller. This suggests that, on average, five purchases were added for each purchaser added due to ad spending. Our estimates also imply an average incremental cost per customer of about $55.

Column (4) of Table 4 shows the effects on the total number of recorded conversions, i.e., also including conversions not specifically related to sales. Advertisers had 0.40 more conversions for every dollar spent. Since the zero spend baseline corresponds to zero conversions, these are all incremental. Comparing these results against those in Columns (2) and (3), it is clear that ad spending increased activity on advertisers' websites more generally despite all campaigns in our experiment being optimized to boost sales.

Finally, in the Appendix we undertake several robustness checks to ensure the quality of our benchmark results. These all confirm our conclusions, and we present them in the Appendix.

5 Learning and Sophistication

Having established our benchmark results on the overall returns to advertising, we turn our attention to a main contribution of our paper: illustrating the concept of *learning by doing*. We do this by exploiting the size of our experiment and the rich heterogeneity across observable advertiser and campaign characteristics to explore which advertisers learn, and how.

5.1 Initial Evidence

Our motivating evidence for learning by doing begins with a simple illustration of how advertiser experience is associated with higher returns on ad spend dollars. We use advertisers' historical ad spending—the total ad spending prior to the week of the experiment—as a proxy measure for experience. All else equal, advertisers who have historically spent more on advertising are likely to have greater advertising experience.

To estimate this, we split our sample to separate advertisers below and above the median level of historical ad spending and estimate equation (1) for each subsample. These results are reported in Table 5. Panels A and B report the results for above and below median subsamples, respectively. While the revenue per ad spend dollar is about \$1.50 for advertisers with below median historical ad spend, the estimate is well over \$3 for advertisers with above median historical ad spend; these two figures are statistically different at the 99% level. This stark contrast illustrates how greater advertising experience is likely to lead to better outcomes for advertisers.

We conduct a battery of robustness checks to establish the validity of our motivating results. First, we turn to three alternative measures of advertiser experience. Specifically, we consider three parameters: (i) the number of ads run on Meta by the advertiser, (ii) the number of ad campaigns run on Meta by the advertiser, and (iii) the number of days since the advertiser first created an ad on Meta. For each of these measures of experience, we repeat the above exercise: we split the sample by below and above median experience and estimate equation (1) for each subsample separately. Columns (2) - (4) of Table 5 reports these results. In each case, we note a difference between less-experienced and more-experienced advertisers, where greater experience is associated with higher returns on advertising.[22]

Second, to ensure that our effects are not driven by advertisers who only advertise sporadically, rather than consistently over a longer period, we repeat our above exercises with advertisers who created their first ad at least 30, 60, 90, and 180 days prior to the start of the experiment, respectively. We find that the estimates are virtually unchanged when imposing this additional restriction, giving us confidence that our effects are not being driven by certain advertisers who may soon stop advertising. These results are presented in Appendix Tables 7.1 - 7.4.

Third, we split the sample of advertisers by quintiles, rather than just the median, of experience and estimate equation (1) for the observations that fall within each quintile. Figure 2 shows the relationship between the effect of ad spending on revenues, as measured by our experiment, and each quintile of historical ad spending, our primary measure of advertiser experience. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval. The data highlights a clear positive trend, consistent again with more experienced advertisers obtaining greater effectiveness of their ads on Facebook and Instagram. Advertisers in the highest quintile had estimated effects higher than \$3 per dollar spent and significantly higher than that of the middle three quintiles (with monotone effects at around \$2); advertisers in the lowest quintile had estimated effects less than \$1 and significantly lower than the top quintile. We

[22] The ad spend coefficient is statistically different at the 99% level for two of three variables capturing experience.

also further measure these effects flexibly with deciles, rather than quintiles, of historical ad spend. These give us consistent estimates as well and are reported in Appendix Figures A2 - A4.

We also note that it is possible for large advertisers to be new to Facebook or Instagram and to immediately spend a relatively large amount on their advertising campaign. As a result, total historical spend may not capture advertiser learning since a smaller advertiser may have spent much more time on Facebook or Instagram, but still have spent cumulatively less than the large newcomer. The correlation coefficient between historical ad spending quintiles and advertiser age quintiles is 0.63. The correlation coefficient between historical ad spending quintiles and campaign age quintiles is 0.30. As further checks, we examine the relationship between campaign age and the effect of ad spending in Figure 3, and the relationship between advertiser age and the effect of ad spending in Figure 4. Each figure shows a positive relationship, consistent with advertisers learning how to improve their advertising effectiveness over time.

Finally, our results remain qualitatively robust if we conduct the exercise for the largest spending and all other advertisers separately. That is, we first split the sample by whether an advertiser is in the top decile of historical ad spend or not. Then, within each subsample, we further split by whether advertisers are below, or above median historical ad spend. These results are also documented in Appendix Figure A5. We observe that re-estimating equation (1) on these subsamples preserve our results. In fact, the difference between above and below median experience is starker for the smaller spending advertisers. We interpret this as suggestive evidence that the returns to experience are larger for smaller spenders.

5.2 Establishing Advertiser Learning

Improved advertising performance over time, or with more experience, may not necessarily be a consequence of advertiser learning by doing. It is possible that users are learning more about the company, or it is possible that Meta's algorithm is learning to improve performance over time. In this section, we distinguish advertiser learning by doing in our experiment from the two alternative explanations for higher returns to spending with experience. To do this, we focus on a measure of advertiser engagement on Facebook and Instagram.

In particular, we look at whether the advertiser has updated their campaign(s) in the past.[23] We are interested in this measure as it is something controlled solely by the advertiser and cannot be affected by neither the advertisers' target users' actions nor Meta's ad algorithms. In other words, this is the kind of advertiser behavior that is in line with Arrow (1962): "*Learning can only take place through the attempt to solve a problem and therefore only takes place during activity.*" We divide our observations into campaigns that have made: (i) no updates, (ii) below the median number of updates, conditional on having made any updates, and (iii) above the median number of updates, conditional on having made any updates.[24] While the first isolates potential algorithmic improvements and user familiarity effects, the others indicate advertiser actions, and potentially learning.

Panels (a)-(c) in Figure 5 shows the relationship between campaign age and advertising returns for campaigns in each of the three buckets. We observe that for campaigns with above the median number of updates, there is a clear positive relationship, with statistically distinguishable point estimates. In contrast, there is no such clear relationship for campaigns with few or no updates. While there is an initial upward trend, it is not sustained as the campaign age increases.

This suggests that relative to advertiser learning, neither user familiarity nor algorithmic learning (for that ad campaign) have meaningful impacts over time on the effect of ad spending.[25] If this were not the case, we would have expected a sustained increase in the effects of ad spend over time, regardless of whether the advertiser made updates to their campaigns or not. Taken together,

[23] These updates may include a variety of actions such as modifying budget, targeting, and other features of a campaign. In our analysis, updates include the following: changing campaign name, pausing a campaign, updating the number of ad groups or ads (adding or deleting), changing the end date of a campaign, changing the ads category of the campaign, modifying budget (e.g., amount, optimization, etc.), changing buying type, and the change in use of system tools (e.g., targeting expansion).

[24] The median campaign had 3 updates, excluding the initial setup. This relatively low number is due to the nature of the on-platform data, which allows us to identify the distinct times when a campaign was updated. To the extent that the advertiser made multiple updates—such as modifying budgets and updating a new creative—in a single session, we would count that as 1 update.

[25] We also run a secondary test on potential algorithm learning by examining campaigns with the Targeting Expansion feature enabled versus those without. This feature is an opt-in tool offered by Meta requiring a campaign to simply tick a menu box and which then expands the campaign targeting to a broader algorithmically-suggested audience. Here we see lower returns (via our primary regression) for campaigns with the feature enabled versus those without the feature enabled. If algorithm learning were the primary determinant or even a relatively larger determinant of improved performance, we would expect the opposite relationship.

these estimates support our hypothesis that advertisers learn by doing, and that allows them to get higher returns on their ad spend over time.

We note that it may be possible that some campaigns were not updated since the advertiser already learned how to effectively advertise its products based on its experience updating other campaigns. To accommodate this possibility, panels (a) and (b) in Figure 6 repeat a similar analysis as Figure 5 using advertiser age instead of campaign age. We split the sample by advertisers that have, across all their campaigns in our experiment, made: i) no updates, ii) below the median number of updates, conditional on having made any updates, and iii) above the median number of updates, conditional on having made any updates. Figure 6 shows relationships consistent with those found in Figure 5: there is a positive relationship for advertisers that have updated their campaigns, and none for those that have never updated a campaign.

5.3 Advertiser Sophistication

The above results demonstrate improved advertising performance over time, as advertiser experience increases. Further, we establish that advertisers' active engagement with their campaigns is a primary driver of those improvements. While this implies that advertiser learning is a major component of improved advertising performance over time, the mechanisms surrounding which key features allow advertisers to learn remains unclear.

We posit that there are complementarities between advertiser learning and having access to better quality data. In other words, better data allows advertisers who wish to tweak and learn from changes to their campaigns to reach a level of sophistication, enabling them to get higher marginal returns from their learning over time. Meta's Ads Manager tool provides information to all advertisers to track the performance of their ad campaigns, as well as several actions they can take to inform their overall understanding of Meta's advertising environment and the overall performance of their campaigns. We focus on two of the common actions taken by advertisers with higher quality of underlying data for additional analysis.[26]

[26] We also test the validity of these actions as metrics for sophistication through a Principal Components Analysis. The results of that exercise confirm these metrics as statistical representations of sophistication. We discuss this analysis further in the Appendix.

First, advertisers can choose to implement Pixel tracking for several events, often beyond the specific goals their campaigns are optimized to achieve. This provides advertisers with rich information on which users take specific actions on their webpages, which they can then use to retarget or learn more about potential customer characteristics. For example, suppose an advertiser observes that some users have recorded an "Add to cart" event multiple times, signaling purchase intent, but without completing the purchase. Advertisers can then update the current campaign to target users who resemble similar characteristics. Similarly, they can include such users in future campaigns for related products.

Second, advertisers can choose to install the Conversions API (CAPI; also referred to as the Server Side API or S2S), a tool developed by Meta that mitigates the risks of browsers affecting Pixel data. CAPI allows advertisers to send events directly from their servers to Meta's Ads Manager tool. CAPI offers some unique benefits over regular pixels that can provide advertisers with better underlying data. For example, CAPI is able to capture deeper-funnel events that some Pixels may not accurately capture, such as if users are redirected to a partner's website to complete a payment transaction. Similarly, CAPI can record events that Pixels may miss due to momentary connectivity issues. Finally, CAPI can be easily set up by advertisers and does not require any significant technical implementation.

We first explore some descriptive statistics on how campaign update behavior of sophisticated advertisers—those who take the steps to obtain better quality data—differs from other advertisers. We observe that advertisers who track more than 1 Pixel event—the median across advertisers in our experiment—across their Meta ad campaigns, make, on average, 23% more updates across their campaigns. Similarly, advertisers with CAPI integrations make, on average, 28% more updates across their campaigns. This suggests that there is a positive association between advertisers' decisions to get access to better quality data, and choices of making updates to their ad campaigns. To examine whether sophisticated advertisers are also better learners, we repeat the exercise from Section 5.2 with high or low sophistication as the defining split of the sample. We hypothesize that if sophistication drives advertiser learning, we expect to see ad campaigns experiencing increasing effects of ad spend over time for sophisticated, but not unsophisticated, advertisers' ad campaigns.

Figures 7 and 8 confirm our hypothesis and present these results. Like campaigns with above median number of updates, there is a clearly increasing relationship between campaign age and the effect of ad spend on revenues for sophisticated advertisers, where sophistication is measured as either the advertiser having a CAPI integration, or the advertiser having tracked more than 1 pixel event across its ad campaigns[27]. Similarly, like campaigns with few updates, less sophisticated advertisers show no relationship between campaign age and returns on their ad spend dollars, over a longer period of time. These results highlight that sophistication, whereby advertisers obtain greater quality underlying data on their ad campaigns, is a mechanism that allows advertisers to learn from and improve on the performance of their ad campaigns.

5.4 Additional Evidence: Ad Aesthetic Score and Advertiser Experience

Our discussion on advertiser experience and learning has focused on the key outcome of our analysis—the effect of ad spending on advertiser revenues. In this section, we briefly turn our attention to ad aesthetic scores, a metric that is determined purely by the overall visual appeal of ads—such as whether the image is of high resolution and not pixelated—and unrelated to any actions users take in response to the ads.[28] That is, only actions taken by the advertiser influence the ad aesthetic score.

We describe results from two simple tests related to aesthetic scores. First, we test whether ad aesthetic scores increase over time, especially for relatively new advertisers. If advertisers are learning what works over time, we would expect the visual appeal of their ads to get better over time. We separate campaigns into buckets constituting 30-day increments of the advertiser's age, including all advertisers in our experiment who are up to a year old. Using a non-parametric Mann-Kendall test (Mann, 1945; Kendall, 1975), we find evidence that there is a monotonic increase in ad aesthetic scores by advertiser age over the first year of advertising. Figure A6 in the Appendix illustrates the trends in these mean aesthetic scores by advertiser age buckets.

[27] These results are not intended to identify causal effects of sophistication tools such as CAPI integrations. Rather, we use cross-sectional variation in advertiser age and sophistication to illustrate how sophisticated advertisers are likely to learn over time, and therefore improve their ROAS.

[28] Aesthetic scores are an internal measure of the quality of images in ads. It is an index metric that summarizes various tangible, visual aspects of ad creatives, including blurriness, balance, and numerous other aspects.

Second, we estimate differences in mean aesthetic scores across advertiser dimensions which we identified in previous sections as capturing experience, learning, and sophistication. Using simple t-tests, we show that ad aesthetic scores are increasing across each of these dimensions. We present these results in Table 6.

These simple statistics highlight that ad aesthetic scores improve as advertisers gain more advertising experience. While the downstream implications of these improvements in terms of sales and revenues are beyond the scope of this paper, and less easy to interpret, we take this as additional evidence of advertisers learning and improving their advertising behavior over time.

5.5 Learning, Sophistication, and Survival

One of the most important and robust findings in the productivity literature is the positive correlation between productivity and survival (e.g., Syverson, 2011). Naturally, more productive companies will be better able to weather downturns in demand for their products, or entry by more productive entrants into their markets. If mastering how to best use their advertising dollars is one way in which firms become more profitable, and hence more productive, then one would expect a positive correlation between advertising effectiveness and survival. More importantly, if traits such as a knack for learning and being sophisticated in advertising extend to other aspects of running a successful establishment, then we'd expect those who are more engaged in learning and exhibit higher degrees of sophistication to survive longer.

We don't have data on whether an establishment in our sample survives as a business entity because we do not link our data to other sources that document a firm's status. We do, however, see whether any advertiser in our experiment continues to advertise after the experimental period. As such, we turn to investigate the propensity of each of the four groups we identify above (high/low learners and high/low levels of sophistication) to continue to advertise on Facebook and Instagram 8 months after the experiment was conducted.[29]

Let $0 \leq p_{jk} \leq 1$ be the proportion of advertisers in group jk in our sample who continue to actively advertise eight months after our weeklong experiment, where $j \in \{L, H\}$ denotes whether they are below (L) or above (H) the median in terms of learning activity and $k \in \{L, H\}$ denotes

[29] The patterns we find are qualitatively the same if we consider shorter periods such as 3 or 6 months.

whether they are below/above the median in terms of data sophistication. This gives us four "survival" probabilities based on learning and sophistication, where we hypothesize that p_{HH} has the highest value of the four, p_{LL} the lowest, and the other two are somewhere in between the extreme values

We turn to estimate these four probabilities and then divide each of them by p_{HH} so as to normalize the probabilities in relation to p_{HH} in order to preserve sensitive company information. The estimates [and 95% confidence intervals] are as follows: $p_{HH} = 1$, $p_{HL} = 0.5980\ [0.5942, 0.6018]$, $p_{LH} = 0.8212\ [0.8204, 0.8222]$, and $p_{LL} = 0.3069\ [0.3043, 0.3094]$. The estimates clearly show that survival as an advertiser is strongly correlated with whether the advertiser is a more engaged learner and a more sophisticated decision maker. The differences suggest that sophistication plays the more important role, as evidenced by high levels of sophistication with low levels of learning activity producing a higher probability of survival than vice versa (low sophistication and high learning activity).

As mentioned earlier, we do not have actual survival data, but conjecture that those who are below median in both learning and sophistication would be more likely to go out of business regardless of the return on advertising they achieved during the experiment. Namely, consider two advertisers, A and B who are in the LL group, where A had below median returns on ads and B above median returns. As they are neither active learners, nor sophisticated decision makers, we'd expect them to go out of business regardless of the performance during the experiment – any variation for businesses in the LL group is, shall we say, dumb luck. Hence, we'd expect them to be equally likely to go out of business and stop advertising. This would not be true for the HH group. If advertisers C and D are in the HH group, where C had below median returns on ads and D above median returns, then their sophistication suggests that the platform may be a better advertising tool for D than for C. As a consequence, we expect D to be more likely to continue advertising on the platform.

We test for the above scenario by examining the differences in survival rates based on return on ad spend for above median versus below median advertisers for each of our four cells of learning activity and sophistication (i.e. we split the HH, HL, LH, and LL groups each into above and below median for return on ad spend). These results support our hypotheses by showing statistically significant differences in survival rates for the HH, HL, and LH groups—indicating that for sophisticated decision makers or active learners (or both), advertising returns are related to survival. In contrast, there is no statistically significant difference in survival rates between advertisers with above and below median advertising returns in the LL group. Therefore, only in the LL do we see that survival rate is independent of advertising returns. Tables 7 and 8 summarize these results and those in the preceding paragraph.

6 Discussion

We have provided evidence that first, establishments' digital advertising expenditures lead to substantial revenue generation on two major social media platforms; second, there is significant variation in the returns to advertising across establishments; third, the behavior of some advertisers is consistent with patterns of learning by doing—those who tweak their campaigns more often tend to realize significantly higher returns on advertising; fourth, advertisers who we consider to be more sophisticated—defined as collecting more data that support a richer analysis of their users' behavior—see even higher returns to their learning activities; last, advertisers who are more sophisticated and more active learners continue to advertise for longer, suggestive of the survival patterns described in the literature on productivity and survival.

Though difficult to document due to data scarcity to date, it should not be surprising that learning how to advertise is important for performance, and that using rich data to support learning leads to more effective outcomes. The prior empirical literature in learning by doing (e.g., Benkard, 2000; Hendel and Spiegel, 2014; and Levitt, List, and Syverson, 2013) focused primarily on production or revenue productivity as these were available data. Another non-production-process activity that may require some learning and sophistication for success is pricing. Huang, Ellickson and Lovett (2022) study the 2012 privatization of off-premises liquor sales in Washington State and find that establishments learn to set more profitable prices, which increasingly reflect demand fundamentals over time.

As for the use of data in learning, Hanna, Mullainathan, and Schwartzstein (2014) use a field experiment to show that even when data are available, failure to notice it results in a lack of learning. Hence, data availability is necessary but not sufficient for learning. We show that advertisers who actively seek to acquire more data do indeed learn more from their experimentation. This also relates to two recent papers that explore using experimentation as a learning activity in firms. Koning, Hasan, and Chatterji (2022) provide evidence on how digital experimentation by high-technology start-ups improves performance by 30%–100% after a year of use using data that tracks their growth, technology use, and products, yet relatively few firms adopt A/B testing. Runge, Geinitz, and Ejdemyr (2022) explore behavior of firms advertising on Facebook and show that a minority of firms engage in A/B testing with their ads, and those who experiment in a given year see higher returns in the following year. However, as the authors explain, the firms they analyze come from a highly selected sample that receive support from Facebook, and there is no experimental variation in ads or ad spend, making a causal interpretation challenging.

There is ample evidence for the growing role of marketing and sales personnel in U.S. companies (Bronnenberg, Dubé, and Syverson, 2022). But due to the scarcity of high-quality data, the economics literature has largely ignored the role of customer-facing activities in shaping how businesses organize, and how markets, firm productivity, and macroeconomic growth are affected by these customer-facing activities. Importantly, as Bronnenberg, Dubé, and Syverson (2022) argue, "*[t]he productive benefits of brand capital and the human capital associated with marketing expertise are not currently considered in standard productivity measurement. This omission would likely under-state productivity at the time marketing investments are made, while over-stating subsequent productivity when marketing investments pay off through increased consumer demand*." We humbly believe that our paper takes a first step in providing evidence that there is significant variation in how well businesses are able to use advertising, and how propensities to learn and use more sophisticated data help the companies who actively engage in these behaviors. As more and more data are recorded administratively in firms, from HR and production data to pricing and advertising data, it may be possible to soon enhance the standard capital-labor models used in the literature to take account for customer facing and other internal activities that help firms become more productive and profitable.

Finally, there are also broader policy implications to these findings. Because certain types of data are important to advertiser learning, limiting the availability of said data would also tend to limit advertisers' ability to learn and achieve higher returns. In the case of the experimental analysis in this paper, offsite signals—the Pixel data that advertisers share with Facebook and Instagram—are of particular importance in driving improved returns. Government regulations that restrict the sharing of these data, as well as technical interventions that functionally prevent it (e.g., Apple's App Tracking Transparency program), could lead to broader economic harms by preventing businesses from learning and improving their fortunes. While this paper does not seek to identify the socially optimal balance of such regulations and restrictions, it does underscore an important cost that should be considered carefully as decisions are undertaken.

References

Abowd, John M., John Haltiwanger, Ron Jarmin, Julia Lane, Paul Lengermann, Kristin McCue, Kevin McKinney, and Kristin Sandusky. 2005. "The Relation among Human Capital, Productivity, and Market Value: Building Up from Micro Evidence." In *Measuring Capital in the New Economy*, ed. Carol Corrado, John Haltiwanger, and Daniel Sichel, 153–98.

Arrow, Kenneth J. "The Economic Implications of Learning by Doing*." The Review of Economic Studies*, vol. 29, no. 3, 1962, pp. 155–73. JSTOR, https://doi.org/10.2307/2295952.

Backus, Matthew, Thomas Blake, Jett Pettus, and Steven Tadelis. (2023) "Communication, Learning, and Bargaining Breakdown: An Empirical Analysis," NBER working paper #w27984.

Benkard, C., Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review*, 90 (4): 1034-1054.

Blake, Thomas, Chris Nosko, and Steven Tadelis. 2015. "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment." *Econometrica*, 83(1): 155-174.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices across Firms and Countries." *Quarterly Journal of Economics*, 122(4): 1351–1408.

Bronnenberg, Bart J., Jean-Pierre Dubé, and Chad Syverson. 2022. "Marketing Investment and Intangible Brand Capital." *Journal of Economic Perspectives*, 36 (3): 53-74.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, 62(1): 53–82.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1): 394–425.

Doraszelski, Ulrich, Gregory Lewis, and Ariel Pakes. 2018 "Just Starting Out: Learning and Price Competition in a New Market," *American Economic Review*, 108 (3), 565–615.

Gordon, Brett R., Robert Moakler, and Florian Zettelmeyer. 2022. "Close Enough? A Large-Scale Exploration of Non-Experimental Approaches to Advertising Measurement." arXiv, https://doi.org/10.48550/arXiv.2201.07055

Gordon, Brett R., Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. 2019. "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook." *Marketing Science* 38, no. 2: 193–364.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence from a Field Experiment," *The Quarterly Journal of Economics*, 129(3): 1311–1353.

Hendel, Igal, and Yossi Spiegel. 2014. "Small Steps for Workers, a Giant Leap for Productivity." *American Economic Journal: Applied Economics*, 6 (1): 73-90.

Hermle, Johannes and Giorgio Martini. 2022. "Valid and Unobtrusive Measurement of Returns to Advertising through Asymmetric Budget Split." arXiv, https://doi.org/10.48550/arXiv.2207.00206.

Hoban, P. R., & Bucklin, R. E. 2015. "Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment." *Journal of Marketing Research*, 52(3), 375–393. https://doi.org/10.1509/jmr.13.0277.

Hopenhayn, Hugo A. 1992. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, 60(5): 1127–50.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1403–48.

Huang, Yufeng, Paul B. Ellickson, and Mitchell J. Lovett. 2022. "Learning to Set Prices," *Journal of Marketing Research*, 59(2):411–434.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *American Economic Review*, 87(3):291–313.

Johnson, Garrett A. 2022. "Inferno: A Guide to Field Experiments in Online Display Advertising." SSRN Working Paper, http://dx.doi.org/10.2139/ssrn.3581396.

Johnson, Garrett, Lewis, Randall A., and Nubbemeyer, Elmar. 2017. "The Online Display Ad Effectiveness Funnel & Carryover: Lessons from 432 Field Experiments," Available at SSRN: https://ssrn.com/abstract=2701578.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica*, 50(3): 649–70

Kendall, M. G. (1975). *Rank Correlation Methods.* New York, NY: Oxford University Press.

Koning, Rembrand, Sharique Hasan, and Aaron Chatterji. 2022. "Experimentation and Start-up Performance: Evidence from A/B Testing." *Management Science* 68, no. 9: https://doi.org/10.1287/mnsc.2021.4209

Lazear, Edward P. 2000. "Performance Pay and Productivity." *American Economic Review*, 90(5): 1346–61.

Lewis, Randall A. and Justin M. Rao. 2015. "The Unfavorable Economics of Measuring the Returns to Advertising." *The Quarterly Journal of Economics* 130, no. 4: 1941–73. https://doi.org/10.1093/qje/qjv023.

Lewis, Randall A., Justin M. Rao, and David H. Reiley. 2011. "Here, There, and Everywhere: Correlated Online Behaviors Can Lead to Overestimates of the Effects of Advertising." *Proceedings of the 20th ACM International World Wide Web Conference*: 157–66.

Lewis, Randall A. and David H. Reiley. 2014. "Online Ads and Offline Sales: Measuring the Effects of Retail Advertising Via a Controlled Experiment on Yahoo!" *Quantitative Marketing and Economics* 12, no. 3: 235–66.

Levitt, Steven D., John A. List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Plant" *Journal of Political Economy*, 121(4): 643-681.

Li, Y., & Xie, Y. (2020). Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement. *Journal of Marketing Research*, 57(1), 1–19. https://doi.org/10.1177/0022243719881113

Lundberg, Erik. *Produktivitet och rädntabilitet*, Stockholm: P. A. Norstedt and Söner, 1961.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica* 13, 245–259. doi: 10.2307/1907187

Meta, Inc. n.d. Business Help Center. Accessed February 14, 2023. Available at: https://www.facebook.com/business/help.

Meta, Inc. 2020. "God Questions, Real Answers: How Does Facebook Use Machine Learning to Deliver Ads." Facebook Business. 11 June 2020. Accessed February 13, 2023. Available at: https://www.facebook.com/business/news/good-questions-real-answers-how-does-facebook-use-machine-learning-to-deliver-ads.

Runge, Julian, Steven Geinitz, and Simon Ejdemyr. 2022. "Experimentation and Performance in Advertising: An Observational Survey of Firm Practices on Facebook." *Expert Systems with Applications* 158: 113554.

Strong, E. K. (1925). *The psychology of selling and advertising*. McGraw-Hill book Company, Incorporated.

Syverson, Chad. 2004a. "Market Structure and Productivity: A Concrete Example." *Journal of Political Economy*, 112(6): 1181–1222.

Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49 (2): 326–365.

Wernerfelt, Nils, Anna Tuchman, Bradley Shapiro, and Robert Moakler. 2022. Estimating the Value of Offsite Data to Advertisers on Meta. SSRN Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4198438.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan. "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features" *Management Science* 68, no. 8 (August 2022): 5644–5666.

Tables and Figures

Table 1: Means and Standard Deviations of Dependent and Independent Variables Across Ad-Eligible and Holdout Groups

	Ad Eligible	Holdout
Revenues ($)	181.25	29.70
	(1663.28)	(618.89)
Purchases	3.84	0.36
	(134.26)	(15.45)
Purchasers	0.66	0.06
	(16.77)	(1.87)
Total Conversions	21.53	1.55
	(1357.70)	(146.07)
Budget Spend ($)	31.73	0.00
	(137.93)	(0.00)
Group Size	1697.56	203.30
	(12528.90)	(1444.95)

Notes: This table presents means and standard deviations (in parentheses) by experimental group for the main variables used in our analyses.

Table 2: Median Advertiser Characteristics

	Experiment Advertisers	All Active Advertisers
Lifetime Ad Campaigns	19	18
Account Age (Days)	475	447
7 Day Ad Spend ($)	206.80	33.90
7 Day Clicks	529	139
Ad Aesthetic Score	2.345	2.310
Followers Across Business Pages	1,611	1,167

Notes: This table presents median characteristics among advertisers in the experiment versus all active advertisers on Facebook and Instagram during the week of the experiment. "Active" advertisers correspond to those that had non-zero ad spend during the week of the experiment. The conversion numbers correspond to the type of conversion (e.g., views, purchases) the advertiser's ads are optimized for. Aesthetic scores represent the overall visual appeal of advertisements.

Table 3: Mean User Characteristics by Treatment Exposure

	(1) Ad Eligible	(2) Holdout
Age	42	42
	(17.5)	(17.8)
Female (Share)	0.54	0.55
	(0.50)	(0.50)
Facebook Age	3554	3687
	(1712.6)	(1752.4)
Friend Count	520	548
	(746.4)	(736.5)
Subscriber Count	39	41
	(1630.4)	(1547.3)
Subscription Count	291	312
	(725.7)	(703.1)
Active on Mobile in Last 30 Days (Share)	0.90	0.91
	(0.29)	(0.31)
Active on Web in Last 30 Days (Share)	0.24	0.25
	(0.43)	(0.42)
Married (Share)	0.30	0.32
	(0.49)	(0.49)
Single (Share)	0.17	0.17
	(0.44)	(0.44)
Confirmed Email (Share)	0.93	0.93
	(0.25)	(0.26)
Birthday Visible (Share)	0.89	0.9
	(0.31)	(0.32)
Profile Picture Present (Share)	0.94	0.95
	(0.22)	(0.24)

Notes: This table presents means and standard deviations (in parentheses) of on-platform demographics and other characteristics for users in our experiment.

Table 4: Effect of Advertising Spending on Revenues and Other Outcomes

Dependent Var:	(1) Revenues	(2) Purchases	(3) Purchasers	(4) Conversions
Budget Spend	3.3098***	0.1029***	0.0179***	0.3980***
	(0.1233)	(0.0132)	(0.0021)	(0.1096)
Group Size	-0.0063***	0.0002	0.0000	0.0066***
	(0.0010)	(0.0001)	(0.0000)	(0.0021)
Constant	62.3880***	0.3350	0.0350	-1.1422
	(1.7389)	(0.1993)	(0.0333)	(3.0721)
No. of Observations	1,323,760	1,323,760	1,323,760	1,323,760
R^2	0.696	0.613	0.618	0.604

Notes: This table reports the results of estimating equation (1) for revenues, our primary outcome variable, as well as purchases, purchasers, and conversions, which are our secondary outcome variables of interest. An observation is recorded at the ad campaign – treatment group level. There were ad campaigns from 210,133 advertisers in the experiment. In each regression, the dependent variable is the outcome variable corresponding to each column header. Each regression includes ad campaign fixed effects. Group_Size controls for the effect of spending a given dollar amount across a larger (extensive margin) or smaller (intensive margin) group of users; dropping Group_Size has no statistically significant impact on the Budget Spend coefficient at the 95% level. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Table 5: Effect of Advertising Spending on Revenues by Advertiser Experience

	(1) Historical Ad Spend	(2) Num. Ads	(3) Num. Campaigns	(4) Days Since First Ad
Panel A: Below Median				
Budget Spend	1.4944***	1.9499***	2.1970***	2.8040***
	(0.1367)	(0.2375)	(0.2189)	(0.3074)
No. of Observations	381,802	358,828	312,262	636,330
R^2	0.587	0.655	0.651	0.698
Panel B: Above Median				
Budget Spend	3.3193***	3.3545***	3.4153***	3.4185***
	(0.1240)	(0.1273)	(0.1335)	(0.1342)
No. of Observations	941,958	964.932	1,011,498	687.430
R^2	0.693	0.697	0.692	0.695

Notes: This table reports the results of estimating equation (1) for four different measures of advertiser experience. Column (1) reports results for our benchmark measure: the historical ad spending prior to the week of the experiment. Columns (2) - (4) report results for the number of ads, number of campaigns, and number of days since the advertiser first created an ad (a proxy for advertiser age). An observation is recorded at the ad campaign – treatment group level. There were ad campaigns from 210,133 advertisers in the experiment. Panel A reports the results from regressions using the below median values of the experience variables of interest, while Panel B reports the same using above median values. The dependent variable in each panel is the revenue made by ad campaigns. Ad campaign fixed effects are included in each regression. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Table 6: Difference in Average Ad Aesthetic Score Across Binary Dimensions of Advertiser Experience and Sophistication

	(1) Mean Ad Aesthetic Score: Low Sophistication or Experience	(2) Mean Ad Aesthetic Score: High Sophistication or Experience	(3) Difference in Means
Advertiser Age	2.3386	2.3492	0.0106***
	(0.0003)	(0.0003)	(0.0004)
Campaign Updates	2.3350	2.3462	0.0111***
	(0.0005)	(0.0002)	(0.0005)
Pixel Events Tracking	2.3358	2.3522	0.0164***
	(0.0003)	(0.0003)	(0.0004)
CAPI Integration	2.3353	2.3546	0.0193***
	(0.0003)	(0.0003)	(0.0004)
N (Campaigns) = 642,811			

Notes: This table reports the results of conducting simple difference in means t-tests between advertisers with low vs high levels of experience and sophistication. We use indicators for: i) advertiser age, which takes a value of 1 if the advertiser is of above median age, ii) campaign updates, which takes a value of 1 if any update was made to the campaign, iii) pixel events tracking, which takes a value of 1 if the advertisers tracks above the median number of pixel events, and iv) CAPI integration, which takes a value of 1 if the advertiser has a CAPI integration. We use data from 642,811 campaigns as aesthetic scores are unavailable for the remainder of the campaigns.

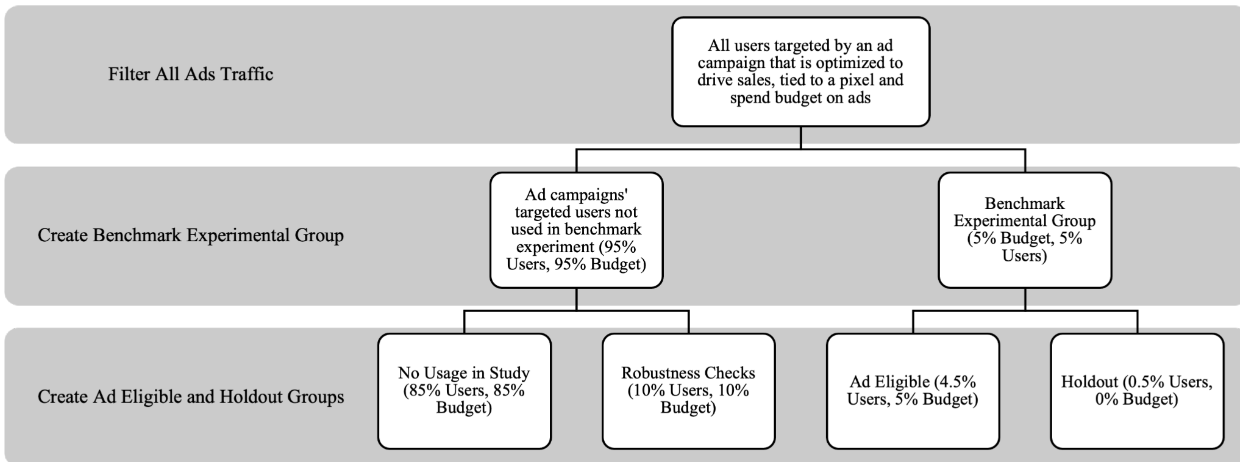
Table 7: Survival Probabilities (Normalized) 8 Months After the Experiment

Learning Activity	Sophistication: L	Sophistication: H
L	0.3069 [0.3043, 0.3094]	0.8212 [0.8204, 0.8222]
H	0.5980 [0.5942, 0.6018]	1.0000

Table 8: Difference in Above vs. Below Median ROA within Each Cell

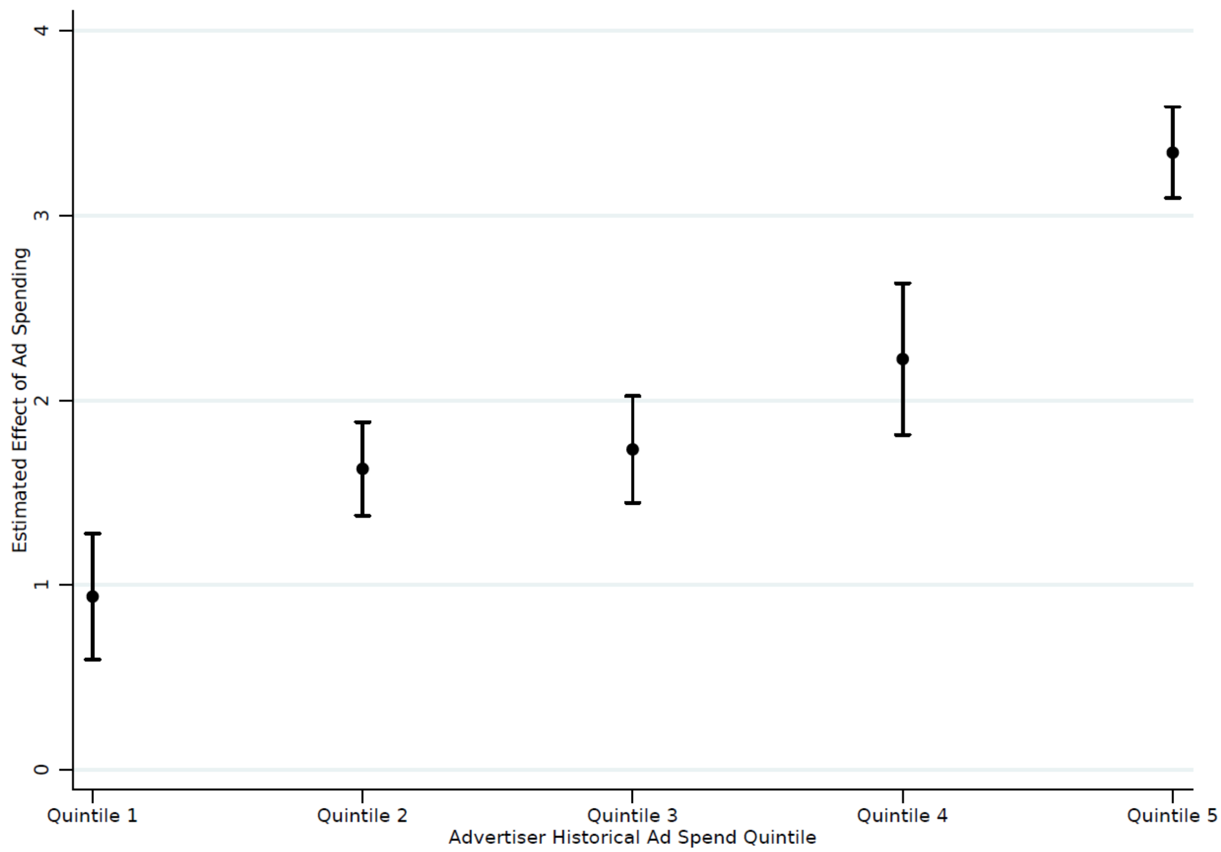
Learning Activity	Sophistication: L	Sophistication: H
L	Difference = -0.0013 p-value = 0.7885	Difference = 0.0813 p-value = 0.0000
H	Difference = 0.0856 p-value = 0.0000	Difference = 0.1271 p-value = 0.0000

Figure 1: Experimental Design for Each Advertising Campaign



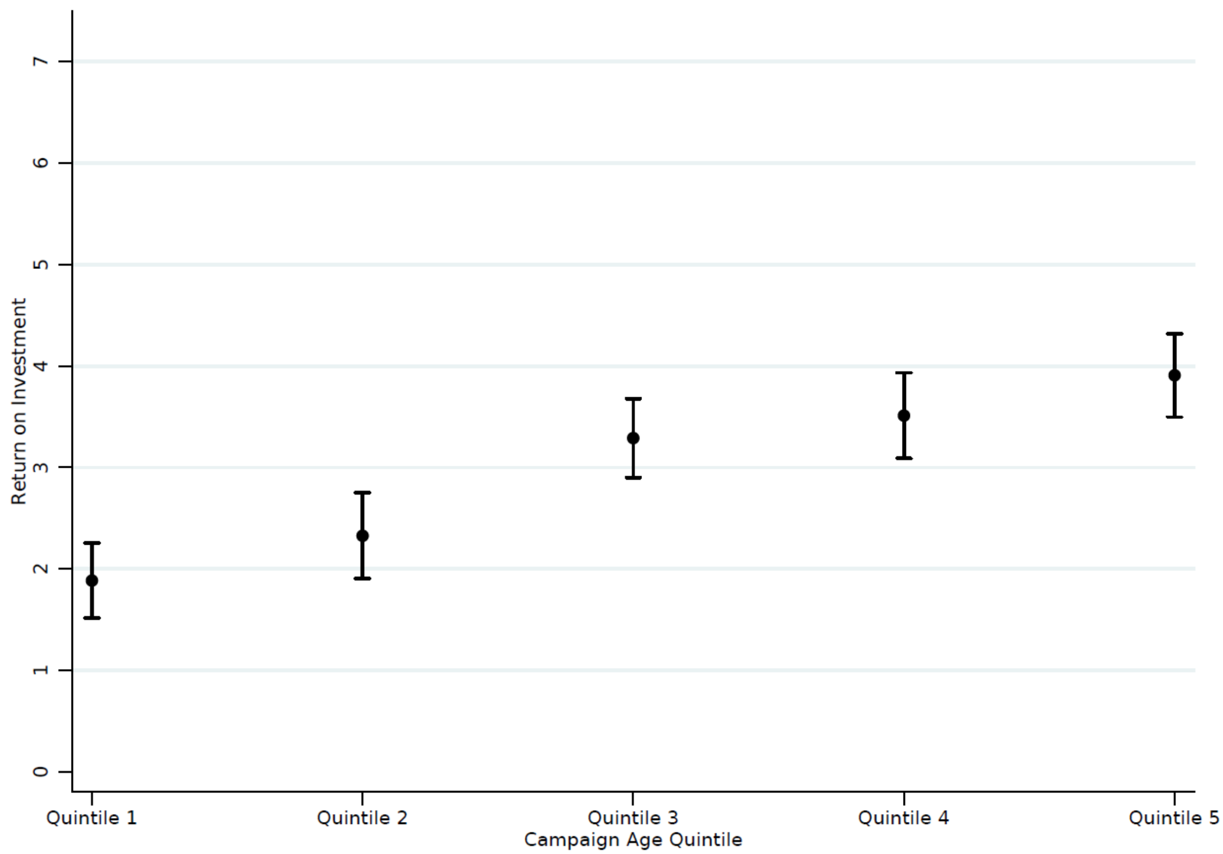
Notes: For each ad campaign in our experiment, 5% of its targeted users were randomly assigned to the experiment; the remaining 95% of targeted users were excluded from the benchmark experiment for that campaign (10% of these were used for robustness checks). The experimental group was further subdivided into a group of Ad-Eligible users, who were given the opportunity to see ads from the campaign, and Holdout users, who were denied the ability to view ads from the campaign. The budgets allocated (internally) by the ad algorithms were approximately balanced between Ad-Eligible and Holdout users, which ensured that Ad-Eligible and Holdout users were comparable. If an ad from an advertiser's campaign won an auction to appear in a user's feed, the ad was sent to the user's feed if that user was in an Ad-Eligible group and the minimum amount necessary to have won the auction was deducted from the advertiser's campaign budget. For Holdout users, the winning ad was replaced with the second-placed ad and the campaign budget that the first-placed advertiser would have spent on Holdout group users was allocated across users in the Ad-Eligible group.

Figure 2: The Effect of Ad Spending by Historical Ad Spend Quintile



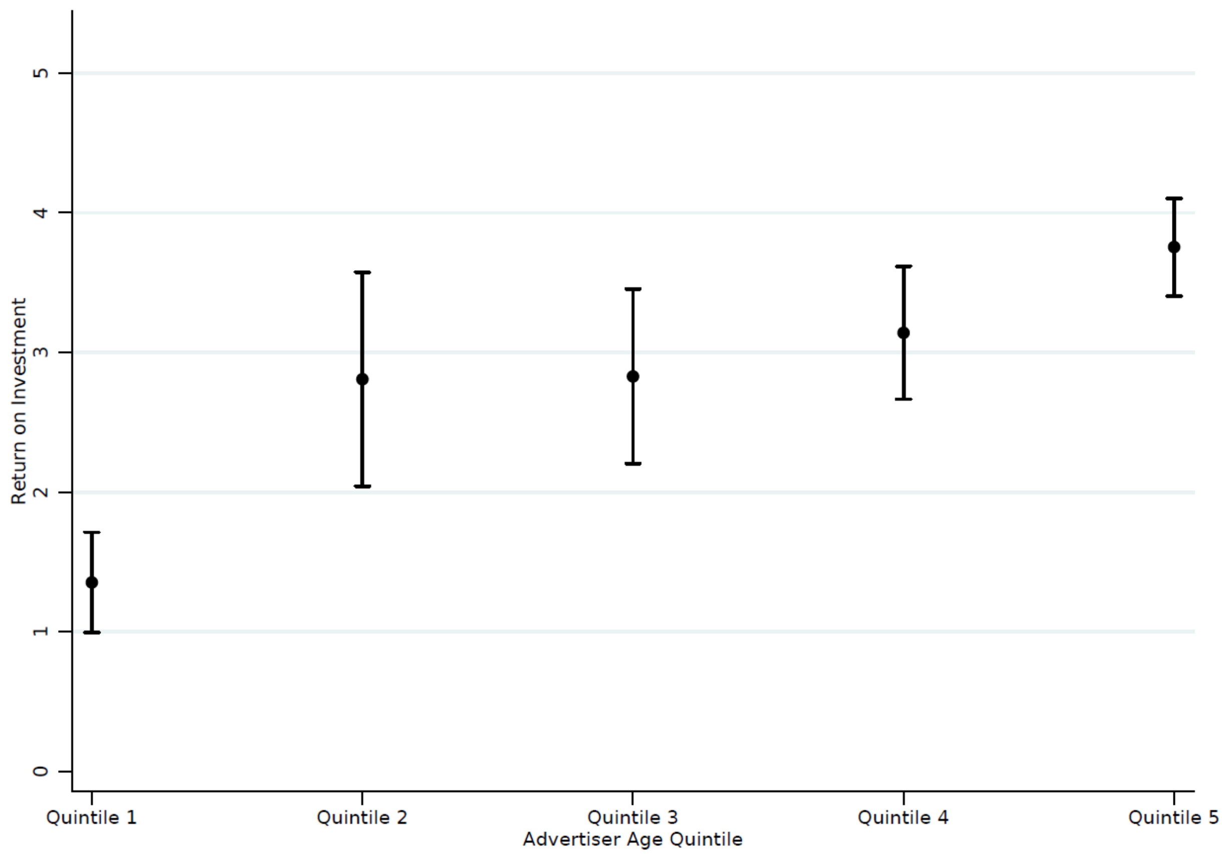
Notes: This figure presents the results of splitting the sample by quintiles of historical ad spending and estimating equation (1) for the observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure 3: The Effect of Ad Spending by Campaign Age Quintile



Notes: This figure presents the results of splitting the sample by quintiles of campaign age and estimating equation (1) for the observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

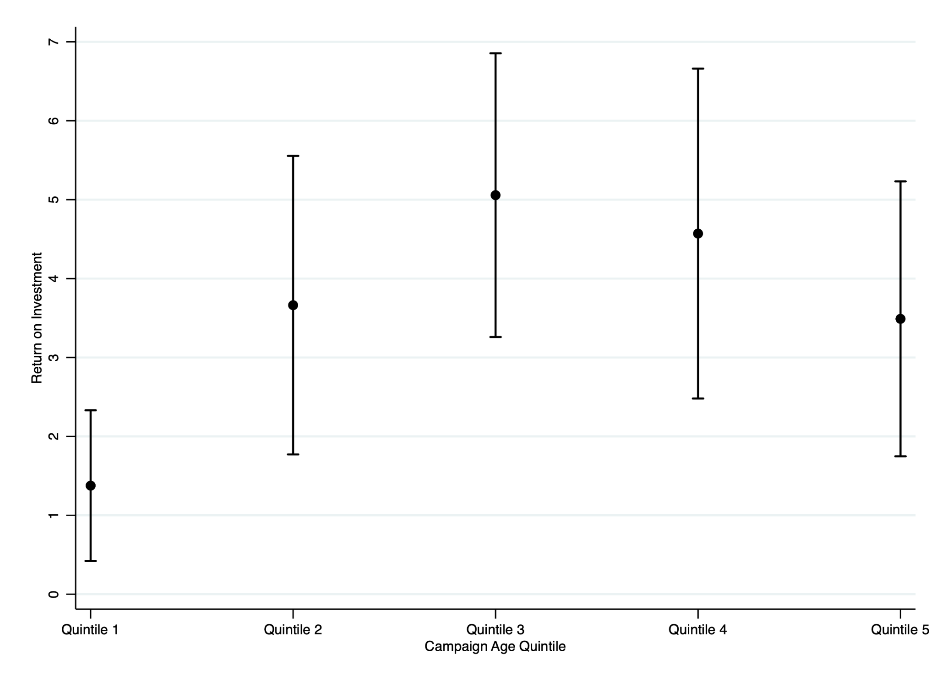
Figure 4: The Effect of Ad Spending by Advertiser Age Quintile



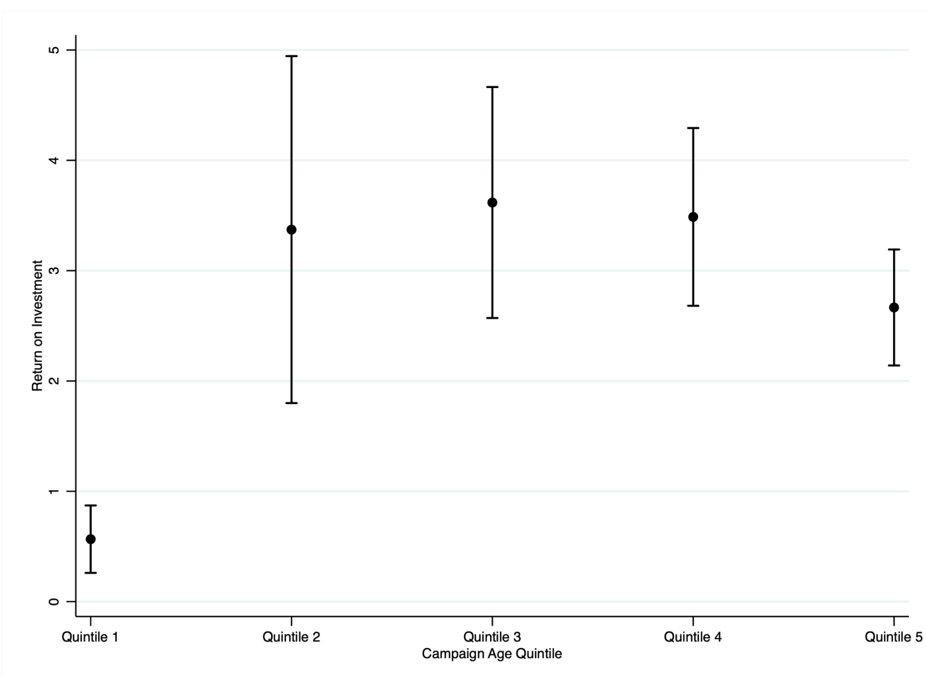
Notes: This figure presents the results of splitting the sample by quintiles of advertiser age and estimating equation (1) for the observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure 5: The Effect of Ad Spending on Revenues by Campaign Age Quintiles and Campaign Update Intensity

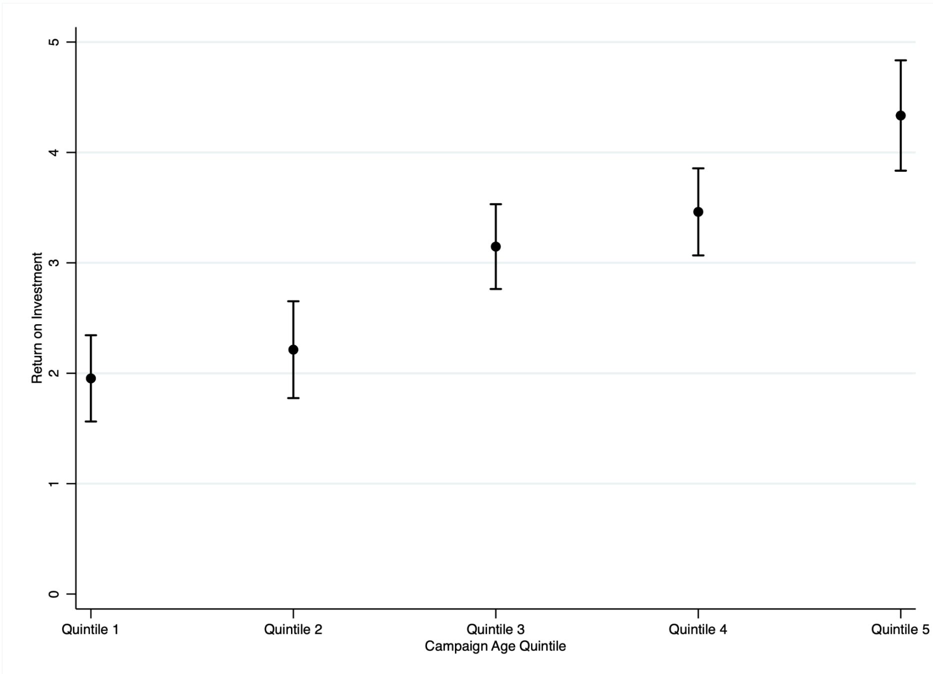
(a) Campaigns That Have Never Been Updated



(b) Campaigns That Have Below Median Updates



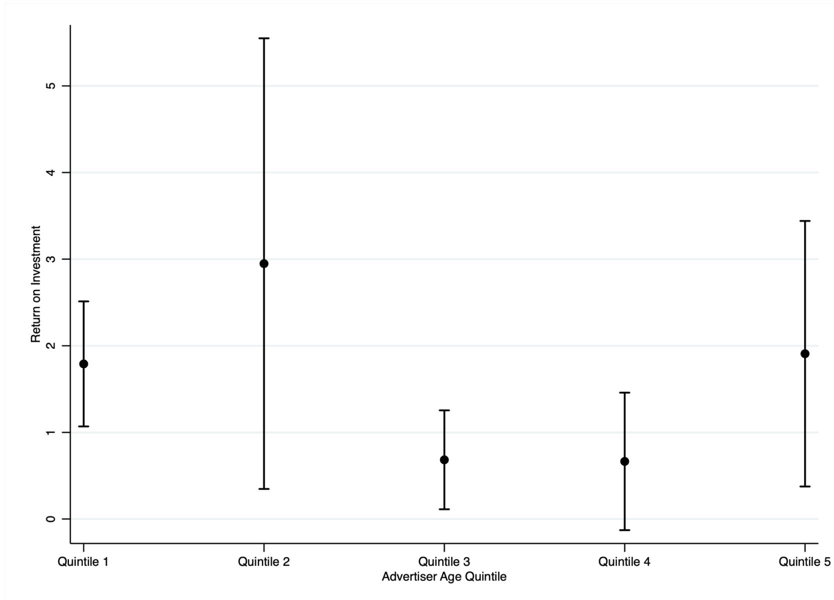
(c) Campaigns That Have Above Median Updates



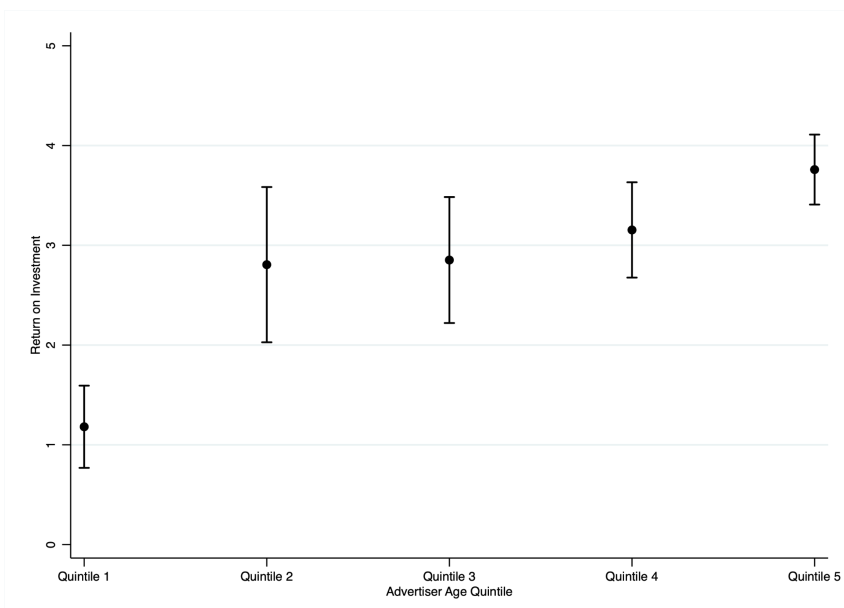
Notes: These figures present the results of first splitting the sample into campaigns that have never been updated, been updated below median number of times, conditional on having any updates, and been updated above the median number of times. For each sample separately, we split the data by quintiles of campaign age and estimate equation (1) for observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure 6: The Effect of Ad Spending on Revenues by Advertiser Age Quintiles and Advertiser Update Behavior

(a) Advertisers That Have Never Updated an Ad Campaign



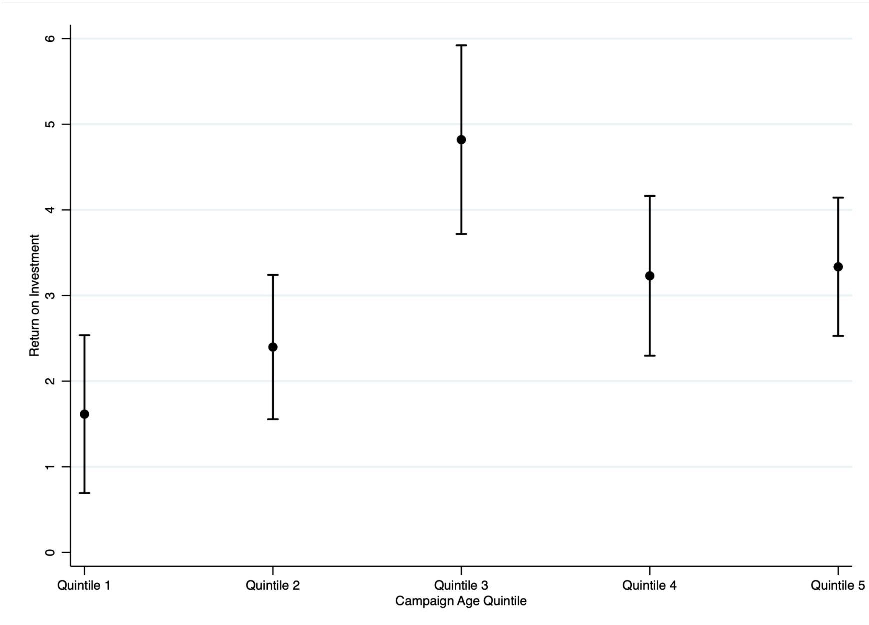
(b) Advertisers That Have Updated an Ad Campaign



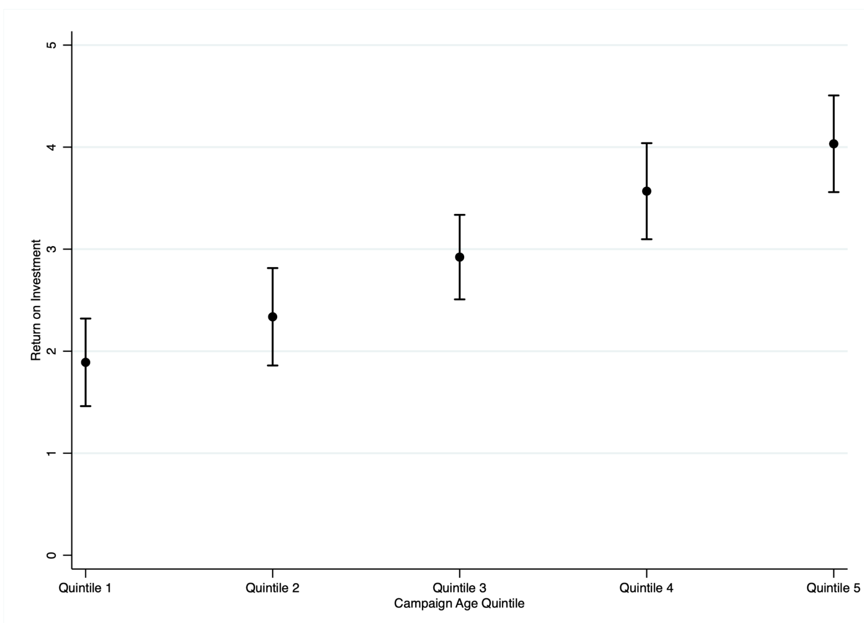
Notes: These figures present the results of first splitting the sample into advertisers that have never updated any of their ad campaigns, and those that have made at least one update across their campaigns. For each sample separately, we split the data by quintiles of advertiser age and estimate equation (1) for observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure 7: The Effect of Ad Spending on Revenues by Campaign Age Quintiles and Pixel Event Tracking Behavior

(a) Campaigns Run by Advertisers That Track Below Median Number of Pixel Events



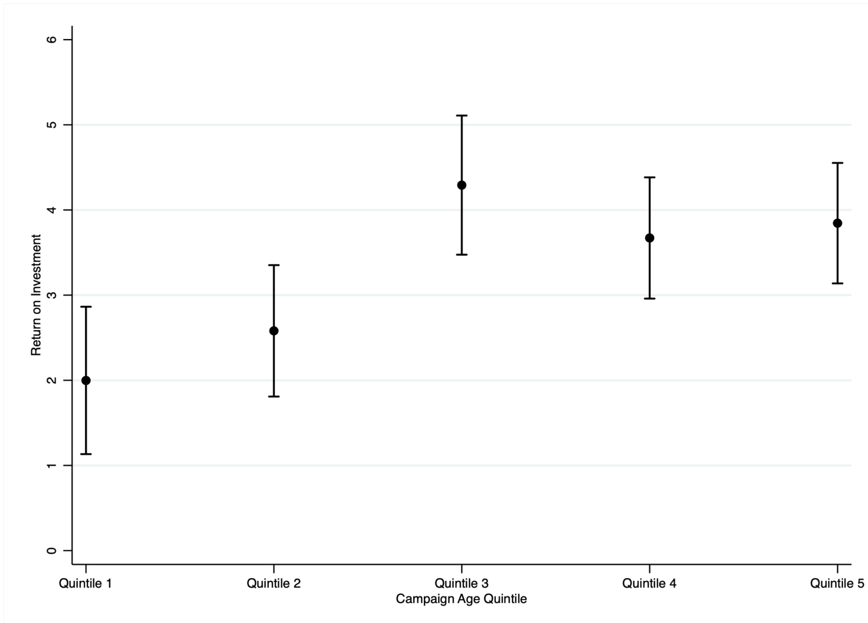
(b) Campaigns Run by Advertisers That Track Above Median Number of Pixel Events



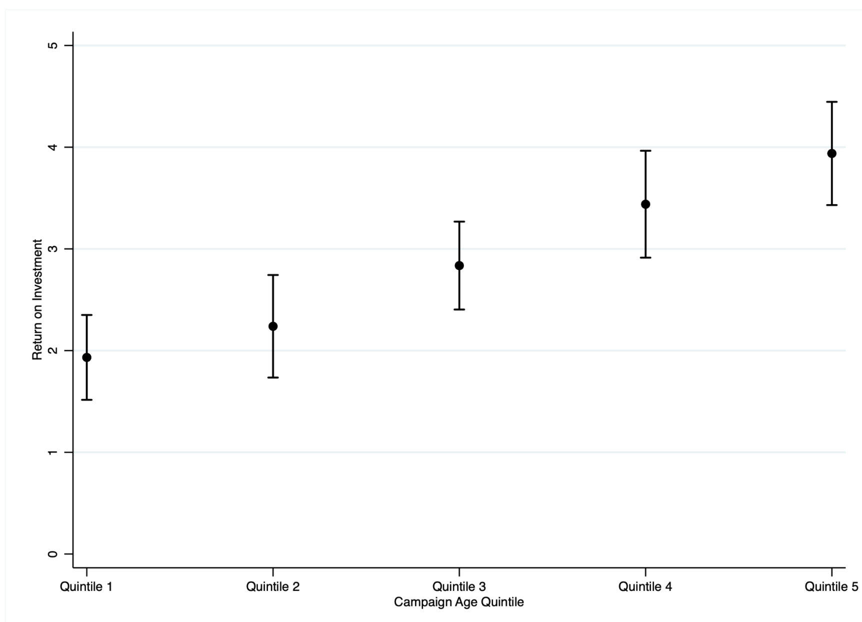
Notes: These figures present the results of first splitting the sample into campaigns run by advertisers that track below and above the median number of Pixel events through their Meta Pixels. For each sample separately, we split the data by quintiles of campaign age and estimate equation (1) for observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure 8: The Effect of Ad Spending on Revenues by Campaign Age Quintiles and CAPI Integration Behavior

(a) Campaigns Run by Advertisers That Do Not Use CAPI Integrations



(b) Campaigns Run by Advertisers That Use CAPI Integrations



Notes: These figures present the results of first splitting the sample into campaigns run by advertisers that do and do not use CAPI integrations. For each sample separately, we split the data by quintiles of campaign age and estimate equation (1) for observations that fall within each quintile. Each bar shows the regression point estimate on Budget

Spend and the corresponding 95% confidence interval. These results use cross-sectional variation in campaign age and use of sophistication tools to document how tools such as CAPI allow advertisers to learn over time and get higher ROAS.

Appendix: Budget Shift Experiment

A. Budget Shift Experiment

We conducted the experiment using pre-existing engineering tools at Meta, which allowed us to withhold some ads that users would normally see and compare outcomes against users who saw their normal ads. These tools also afforded us the opportunity to conduct a “budget shift” component, which tripled our experimental size by adding two additional experimental groups: one with 4% of budget and one with 6% of budget (the former shifting 1% of its budget to the latter). This procedure provided us with a set of ‘built-in’ robustness checks and allowed for more nuance on overall advertiser returns.[30]

For each ad campaign in our experimental universe, the budget shift design consisted of two key components: (i) the creation of three “Budget Shift” groups; and (ii) the creation of “Ad-Eligible” and “Holdout” groups within each Budget Shift group. This leads to the creation of 6 experimental groups for each advertising campaign. This experimental setup is presented in Figure A1.

In total, the budget shift exercise consisted of an additional 10% of each ad campaign’s targeted users, while the remaining 90% were excluded. For each ad campaign, 10% of its targeted users were randomly assigned to one of two approximately equal-sized groups so that there were 5% of users per group. Throughout the Appendix, we refer to these as the “Budget Shift” groups. Each Budget Shift group was also initially assigned 5% of the campaign’s budget. We also note that this is in addition to our primary experimental group, which also had 5% of budget and 5% of users, but without any budget shifting component.

In the first component of the treatment the campaign budget was shifted across groups so that each Budget Shift group was assigned a different budget. The Budget Shift 6% group was assigned 6% of the campaign budget, a 20% increase in the budget relative to our primary experimental group of 5%. Similarly, the Budget Shift 4% group was assigned 4% of the campaign budget, a 20% decrease in the budget relative to our primary experimental group of 5%. These budget shifts occurred within campaigns and ensured that each ad campaign’s overall budget remained unchanged. In addition, the campaign’s budget pacing parameters, targeting criteria, optimization goals, and creatives were unaffected by the budget shifts.

[30] We also note that this setup is similar to that used by Hermle and Martini (2022) and Wernerfelt et al. (2022).

Each Budget Shift group functioned as an independent system within which Meta's ad algorithms optimally allocated the entire group's ad budget across all users in the group. Since all campaigns in our experiment were optimized to boost purchases, the ad algorithms could allocate a positive budget to the subset of users that have a high likelihood of converting after seeing an impression from a specific campaign and allocate zero budget to other users. For a given user, this budget is the Advertiser Bid with which the campaign's ad enters the auction.

Budget Shift groups were further randomly subdivided into a group of "Ad-Eligible users", who were given the opportunity to see ads from the campaign, and "Holdout" users, who were denied the ability to view ads from the campaign. This is the same treatment as our primary experiment and, therefore, each Ad-Eligible comprised 4.5% of the campaign's targeted users, while each Holdout group comprised 0.5% of the campaign's targeted users. In addition to being random, this subdivision was independent of the budget optimization within each Budget Shift group. Therefore, the budgets allocated by the ad algorithms were approximately balanced between Ad-Eligible and Holdout users, which ensured that, within a Budget Shift group, Ad-Eligible and Holdout users were comparable - i.e. a campaign's ad had the same likelihood of winning ad auctions for Ad-Eligible users as it did for Holdout users.

If an ad from an advertiser's campaign won an auction to appear in a user's feed—due to having the highest Total Value in the auction—the ad was sent to the user's feed if that user was in an Ad-Eligible group. In this case, the minimum amount necessary to have won the auction was deducted from the advertiser's campaign budget.[31] If, however, the user was in a Holdout group, the winning ads were withheld after the auction completed and just before ad delivery and replaced with the second-placed ad from the auction. This reflects the counterfactual outcome—the ad winning the auction for a given user—that would have occurred in the absence of the advertiser's campaign.[32] First-placed advertisers were not charged for Holdout users' auctions, as these auctions were effectively "won" by the second-placed advertisers. The campaign budget that the first-placed

[31] As described earlier, this is a modified form of the VCG auction, where the winner pays the minimum amount required to have won the auction.

[32] The counterfactual being reflective of a true $0 ad spend relies on the auction mechanism's stability to the removal of the winning ad (Gordon et al., 2019). That is, the second placed ad should be the same whether the experimental advertiser of interest participated in the auction or not. The assumption here is that other advertisers' strategies remain unchanged in the short run. We believe, echoing Gordon et. al. (2019), that this is a reasonable assumption because campaigns are not pre announced, and the scale of Meta's advertising platforms makes it hard to gauge the campaign scope, strategies, and targeting of other advertisers.

advertiser would have spent on Holdout group users was allocated across users in the Ad-Eligible group (in the same Budget Shift group).[33] Therefore, each advertiser's overall budget was left unchanged.

We provide summaries of all of our benchmark estimations and checks for these two Budget Shift groups alongside our experimental group in Tables A1-A4.

B. Robustness Checks on Benchmark Results

First, we leverage the existing Meta engineering tools (on which the experiment was implemented) to measure returns for two other groups, one where certain ads within a campaign were delivered with 4% of budget and one where they were delivered with 6% of budget (i.e. a 20% budget decrease and 20% budget increase). This exercise allows us to confirm our benchmark results while adding nuance on the overall returns.

As we noted in Section 3.2, due to some users effectively dropping out of the experiment, not all advertisers have both Ads Eligible and Holdout in our experiment. For our first robustness check, we re-estimate equation (1), with revenue as the dependent variable, on the subset of advertisers that have all six experimental groups—which is approximately 94% of our data. Panel (a) of Table A5 shows these results. The point estimates on Budget Spend are very similar to those in Table A3, and are the same when rounded to the nearest cent. For the second robustness check, we re-estimate equation (1) using advertiser fixed effects, instead of the campaign fixed effects we use in our main specification. Panel (b) in Table A5 reports these results. These results are slightly smaller than those in Table A3. However, while advertiser fixed effects may control for similar unobservable factors as campaign fixed effects, advertiser fixed effects will be unable to control for unobservable factors that may differ across campaigns for the same advertiser. For example, a sportswear company may have a tennis shoe campaign targeted at women and simultaneously have

[33] This reallocation of unspent budget may have occurred on the same day or on a later day within the experimental period. To ensure that this reallocation had minimal effect on overall advertiser outcomes, the experimental design involved two additional features. First, each campaign had a daily budget cap, which set the maximum amount of ad spending on a given day. Without daily budget caps, any unspent budget from the Holdout group could have been spent all at once and led to a high number of auction wins and ad impressions that would not have occurred absent the experiment. This could have led to different user behavior relative to a world in which ads are more reasonably spread out over time. Second, as noted in the main text, the reallocation took place within the same Budget Shift group. Were this not the case, the unspent budget could have been reallocated to users in another Budget Shift group and contaminate the experiment.

a basketball shoe campaign targeted at both men and women. Therefore, our preference is still to use campaign fixed effects for our benchmark estimates.

The Meta ad algorithm also ensured that Ad-Eligible and Holdout users were comparable between treatment and control, as previously described. While this helps ensure that our estimated effects are solely due to the change in budget, the Holdout groups may not correspond exactly to what happens outside the experiment, "in-the-wild," when an advertiser spends zero budget on a group of users.[34] To determine whether our baseline results would likely change dramatically if we had an in-the-wild Holdout group, we examine the importance of Meta's ad algorithms on our three Holdout groups. In particular, we re-estimate equation (1) pairing Ad-Eligible groups with alternative Holdout groups. For example, column (2) in panel (a) in Table A6 corresponds to pairing the Ad-Eligible 5% group with the Holdout 4% group. Examining the results by column corresponds to holding an Ad-Eligible group fixed and varying the Holdout group. Our hypothesis is that if each Holdout group in our experiment is an accurate representation of a zero ad spend by the advertiser, comparing them with a given Ad-Eligible group should yield similar results. We show that this is indeed the case. The estimated effects do not differ by much, less than six cents at most. In addition, there does not appear to be a systematic pattern related to budget; the estimated effects for the Ad-Eligible 5% group are higher using either the Holdout 4% group or the Holdout 6% group. These results give us confidence that our estimated effects correspond well to what one would expect in the wild.

Finally, as an additional informative result, we turn back to our 5% budget group and re-estimate equation (1) using the inverse hyperbolic sine (IHS) of our three secondary outcomes of interest, namely purchases, purchasers, and conversions. For the independent variable, we use an indicator that takes a value of 1 for Ad-Eligible groups and 0 for Holdout groups. Our motivation for this regression is to estimate the overall value of digital ads for these purchase relevant outcomes, and

[34] We confirm the fact that there were no unforeseen impacts to normal traffic (apart from the treatment) by comparing ads over delivery diagnostics on an hourly basis for our experimental groups, which is an indicator that other factors (e.g. ad pacing) may be influencing our experimental groups. The average hourly over delivery for all Meta ads traffic during the month of April was 0.261 millicents or 0.32%. The average hourly over delivery for all Meta ads traffic during the period from March through May 2022 was 0.245 millicents or 0.30%. We use these longer time periods as benchmarks. Our experimental groups saw over delivery of 0.32% (4% budget group), 0.30% (5% budget group), and 0.29% (6% budget group), indicating no group experienced above normal ads over delivery. Internal guidance at Meta states that ads over delivery of 0.3% is normal, while 0.5% is high, 0.7% is a threshold flag limit. Additionally, our experimental ads traffic had a 0.992 correlation with overall traffic during the same period and had no instances of movements counter or to a greater extent than general traffic.

interpret these in percentage terms using the IHS measure. Table A7 presents these results. We document that digital advertising on Meta, on average, increases purchases, purchasers, and overall conversions by 25%, 13%, and 73% respectively.

C. Robustness Checks on Advertiser Sophistication Classifications

In Section 5.3.1 above, we identified two key variables that we believe best capture advertiser sophistication, which we show to be a key mechanism driving advertiser learning over time. As an additional and agnostic data-driven approach, we also perform a Principal Component Analysis (PCA) to explore correlations across a number of measurable variables that advertisers can choose, and may capture some aspect of advertiser experience and learning. We identify 12 such variables *a priori*: 1) Days since account's first ad creation (to the experiment start date), 2) Whether an account is managed by a third party, 3) Number of pixel events, 4) Whether an account has a CAPI integration, 5) Whether a campaign uses the automated bid tool to bid for ads, 6) Whether an campaign budget has been updated, 7) Whether a campaign uses a custom audience, 8) Whether a campaign uses a lookalike audience, 9) The number of ads in a campaign, 10) Whether a campaign has been updated in any way, 11) The number of campaign updates, and 12) The account budget for the year leading up to the experiment.

The PCA results indicate two key implications. First, there is no clear Principal Component (or small collection thereof) that can account for a significant proportion of variance among these variables. Four dimensions demonstrate eigenvalues over 1.0 (a common threshold), but these combined account for only 51.4% of cumulative variance. Second, there is no major dimensionality reduction in using Principal Components compared to using the original variables. As many as six variables produce better than expected contributions to the four robust dimensions. These are: 1) the use of a lookalike audience, 2) the use of a custom audience, 3) whether a budget has been updated, 4) whether a campaign has been updated, and 5) the number of pixel events, and 6) number of days from first ad creation to experiment. We further confirm these results via a PCR analysis where we see no improvement in predictive performance for the total number of impressions—one plausible outcome of interest for advertisers—from the usage of Principal Components over the original variables.

Given the lack of a small number of Principal Components in the data, there is not a straightforward metric through which we can capture a "holistic" measure of "learning inputs", and how that may affect advertiser learning over time. Nonetheless, we repeat our empirical exercise from the earlier sections for each of the six aforementioned variables. That is, for each of these variables, we identify its median value, and split the data into below and above the median. Then, for each subsample, we examine the relationship between campaign age quintiles and the effects of ad spending estimates. Our hypothesis, as before, is that if these variables are positively associated with advertiser learning over time, we would expect to see an increasing relationship for advertisers with higher values (above median) of the variable, but not for other advertisers with lower values (below median).

The results for these variables are qualitatively identical to our earlier results, except for the use of a lookalike audience and the use of a custom audience.[35] In other words, the effects of ad spending increase over time for advertisers with higher values of the learning inputs, while this is not the case for advertisers with lower values. We argue that these results reinforce our earlier arguments: advertisers who make updates experience better returns over time and sophistication in terms of use of better data facilitates such learning. Since use of a custom audience or a lookalike audience is, in itself, evidence of neither advertiser engagement with their ad campaigns nor sophistication with access to iteratively better data, we are not surprised that these results are not indicative of advertiser learning.[36]

[35] For the sake of brevity and since they add no additional nuance to our results, we do not present summary figures for these analyses. We instead discuss them briefly above.

[36] Both tools would require that the advertiser knows their custom audience and or lookalike audience to be of a high-quality nature, but we cannot test that assumption prima facie.

Table A1: Means and Standard Deviations of Dependent and Independent Variables

	Budget Shift 4%		Budget Shift 5%		Budget Shift 6%	
	Ad Eligible	Holdout	Ad Eligible	Holdout	Ad Eligible	Holdout
Revenues ($)	178.66	28.02	181.25	29.70	182.81	28.67
	(1648.83)	(593.02)	(1663.28)	(618.89)	(1672.51)	(606.67)
Purchases	3.89	0.35	3.84	0.36	3.78	0.35
	(132.72)	(14.74)	(134.26)	(15.45)	(138.35)	(15.68)
Purchasers	0.67	0.06	0.66	0.06	0.65	0.06
	(16.73)	(1.83)	(16.77)	(1.87)	(17.08)	(1.90)
Total Conversions	21.31	1.50	21.53	1.55	21.69	1.57
	(1344.45)	(141.26)	(1357.70)	(146.07)	(1397.98)	(162.78)
Budget Spend ($)	25.99	0.00	31.73	0.00	37.30	0.00
	(120.14)	(0.00)	(137.93)	(0.00)	(153.87)	(0.00)
Group Size	1703.33	203.56	1697.56	203.30	1699.23	203.73
	(12635.15)	(1455.88)	(12528.90)	(1444.95)	(12467.04)	(1438.27)

Notes: This table presents means and standard deviations (in parentheses) by experimental group for the main variables used in our analyses.

Table A2: Effect of Advertising Spending on Revenues

	(1)	(2)	(3)
	Budget Shift 4%	Budget Shift 5%	Budget Shift 6%
Budget Spend	3.8268***	3.3098***	3.0056***
	(0.1453)	(0.1233)	(0.1102)
Group Size	-0.0067***	-0.0063***	-0.0059***
	(0.0010)	(0.0010)	(0.0010)
Constant	63.3538***	62.3880***	58.4091***
	(1.6487)	(1.7389)	(1.8503)
N	1,328,556	1,323,760	1,318,450
R^2	0.693	0.696	0.692

Notes: This table reports the results of estimating equation (1) for the three Budget Shift groups. *N* refers to the number of observations where each observation is recorded at the ad campaign – treatment group level, separately for each budget shift group. There were ad campaigns from 210,133 advertisers in the experiment. In each regression, the dependent variable is revenues and ad campaign fixed effects are included. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Table A3: Effect of Advertising Spending on Purchases, Purchasers, and Conversions

	(1) Budget Shift 4%	(2) Budget Shift 5%	(3) Budget Shift 6%
(a) Purchases			
Budget Spend	0.1225***	0.1029***	0.0888***
	(0.0154)	(0.0132)	(0.0115)
Group Size	0.0001	0.0002	0.0002
	(0.0001)	(0.0001)	(0.0001)
Constant	0.4311*	0.3350	0.2200
	(0.1878)	(0.1993)	(0.2100)
N	1,328,556	1,323,760	1,318,450
R^2	0.610	0.613	0.611
(b) Purchasers			
Budget Spend	0.0216***	0.0179***	0.0153***
	(0.0025)	(0.0021)	(0.0017)
Group Size	0.0000	0.0000	0.0001*
	(0.0000)	(0.0000)	(0.0000)
Constant	0.0527	0.0350	0.0173
	(0.0324)	(0.0333)	(0.0342)
N	1,328,556	1,323,760	1,318,450
R^2	0.617	0.618	0.617
(c) Conversions			
Budget Spend	0.5134***	0.3980***	0.3191***
	(0.1542)	(0.1096)	(0.0835)
Group Size	0.0059**	0.0066**	0.0072**
	(0.0020)	(0.0021)	(0.0023)
Constant	-0.9934	-1.1422	-1.2820
	(3.1588)	(3.0721)	(3.0477)
N	1,328,556	1,323,760	1,318,450
R^2	0.602	0.604	0.611

Notes: This table reports the results of estimating equation (1) for the three Budget Shift groups. *N* refers to the number of observations where each observation is recorded at the ad campaign – treatment group level, separately for each budget shift group. There were ad campaigns from 210,133 advertisers in the experiment. The dependent variable is different in each panel: panel (a) reports the effects on the number of purchases, panel (b) reports the effects on the number of purchasers, and panel (c) reports the effects on the number of conversions. Ad campaign fixed effects are included in each regression. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * $p<0.05$, ** $p<0.01$, and *** $p<0.001$.

Table A4: Alternative Estimates of Effect of Advertising Spending on Revenues

	(1) Budget Shift 4%	(2) Budget Shift 5%	(3) Budget Shift 6%
(a) All Six Experimental Groups			
Budget Spend	3.8307***	3.3113***	3.0065***
	(0.1454)	(0.1234)	(0.1103)
Group Size	-0.0067***	-0.0063***	-0.0059***
	(0.0010)	(0.0010)	(0.0010)
Constant	67.8626***	66.5226***	62.0850***
	(1.7666)	(1.8560)	(1.9671)
N	1,237,882	1,237,882	1,237,882
R^2	0.693	0.697	0.692
(b) Advertiser Fixed Effects			
Budget Spend	3.7931***	3.3008***	2.9873***
	(0.1460)	(0.1260)	(0.1126)
Group Size	-0.0051***	-0.0048***	-0.0042***
	(0.0010)	(0.0010)	(0.0010)
Constant	60.1229***	58.9128***	55.1418***
	(1.6653)	(1.7740)	(1.8742)
N	1,375,619	1,372,240	1,367,542
R^2	0.464	0.466	0.465

Notes: This table reports the results of estimating variants of equation (1) for the three Budget Shift groups. N refers to the number of observations where each observation is recorded at the ad campaign – treatment group level, separately for each budget shift group. There were ad campaigns from 210,133 advertisers in the experiment. In each regression, the dependent variable is revenues. Panel (a) restricts the sample to advertisers for which we have data on all six experimental groups. Panel (b) uses advertiser fixed effects rather than ad campaign fixed effects. Panel (c) reports estimates from a regression including a quadratic term for Budget Spend. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * $p<0.05$, ** $p<0.01$, and *** $p<0.001$.

Table A5: Sensitivity of Estimated Effects to Holdout Group

	(1) Ad Eligible 4%	(2) Ad Eligible 5%	(3) Ad Eligible 6%
(a) Holdout 4%			
Budget Spend	3.8268***	3.3677***	3.0421***
	(0.1453)	(0.1273)	(0.1139)
Group Size	-0.0067***	-0.0067***	-0.0061***
	(0.0010)	(0.0010)	(0.0010)
Constant	63.3538***	60.8058***	57.3383***
	(1.6487)	(1.7777)	(1.8945)
N	1,328,556	1,326,682	1,325,210
R^2	0.693	0.692	0.692
(b) Holdout 5%			
Budget Spend	3.7603***	3.3098***	2.9921***
	(0.1409)	(0.1233)	(0.1100)
Group Size	-0.0063***	-0.0063***	-0.0058***
	(0.0010)	(0.0010)	(0.0010)
Constant	64.9946***	62.3880***	59.0091***
	(1.6202)	(1.7389)	(1.8526)
N	1,322,172	1,323,760	1,321,962
R^2	0.698	0.696	0.697
(c) Holdout 6%			
Budget Spend	3.7704***	3.3221***	3.0056***
	(0.1431)	(0.1247)	(0.1102)
Group Size	-0.0063***	-0.0064***	-0.0059***
	(0.0009)	(0.0010)	(0.0010)
Constant	64.5194***	61.9567***	58.4091***
	(1.6437)	(1.7631)	(1.8503)
N	1,317,004	1,317,604	1,318,450
R^2	0.693	0.691	0.692

Notes: This table reports the results of estimating variants of equation (1) in which we pair Ad-Eligible groups with different Holdout groups. *N* refers to the number of observations where each observation is recorded at the ad campaign – treatment group level, separately for each budget shift group. There were ad campaigns from 210,133 advertisers in the experiment. For example, column (1) in panel (a) corresponds to pairing the Ad-Eligible 4% group with the Holdout 4% group; the results are identical to column (1) in Table 4. Column (2) in panel (a) in Table A6 corresponds to pairing the Ad-Eligible 5% group with the Holdout 4% group. In each regression, the dependent variable is revenues. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Table A6: Effect of Advertising on IHS of Outcomes

Dependent Var:	(1) IHS (Purchases)	(2) IHS (Purchasers)	(3) IHS (Conversions)
Budget Spend	0.2547***	0.1333***	0.7292
	(0.0037)	(0.0022)	(0.0085)
Group Size	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)
Constant	0.0494***	0.0237***	0.1131***
	(0.0017)	(0.0010)	(0.0039)
N	1,323,850	1,323,850	1,323,850
R^2	0.686	0.708	0.692

Notes: This table reports the results of estimating variants of equation (1) in which we use the inverse hyperbolic sines of purchase relevant outcomes as the dependent variable and an indicator for the Ad-Eligible group as the independent variable. N refers to the number of observations where each observation is recorded at the ad campaign – treatment group level. There were ad campaigns from 210,133 advertisers in the experiment. All regressions include campaign fixed effects. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Table A7.1 - A7.4: Effect of Advertising Spending on Revenues by Advertiser Experience with Sample Restricted to Different Advertiser Age Cutoffs

Table A7.1: Advertisers At Least 30 Days Old

	(1) Historical Ad Spend	(2) Num. Ads	(3) Num. Campaigns	(4) Days Since First Ad
Panel A: Below Median				
Budget Spend	1.7448***	1.7000***	1.8613***	2.7950***
	(0.1355)	(0.1876)	(0.1866)	(0.2392)
N	301,766	296,972	279,534	574,282
R^2	0.577	0.605	0.618	0.707
Panel B: Above Median				
Budget Spend	3.3310***	3.3929***	3.4592***	3.5243***
	(0.1257)	(0.1304)	(0.1359)	(0.1472)
N	781,172	785,966	803,404	508,656
R^2	0.700	0.701	0.702	0.692

Table A7.2: Advertisers At Least 60 Days Old

	(1) Historical Ad Spend	(2) Num. Ads	(3) Num. Campaigns	(4) Days Since First Ad
Panel A: Below Median				
Budget Spend	1.8219***	1.7204***	2.1211***	2.7813***
	(0.1377)	(0.1962)	(0.2435)	(0.2337)
N	280,270	281,854	268,148	541,932
R^2	0.575	0.593	0.612	0.708
Panel B: Above Median				
Budget Spend	3.3558***	3.4132***	3.4593***	3.5773***
	(0.1270)	(0.1314)	(0.1365)	(0.1510)
N	750,036	748,452	762,158	488,374
R^2	0.697	0.702	0.703	0.691

Table A7.3: Advertisers At Least 90 Days Old

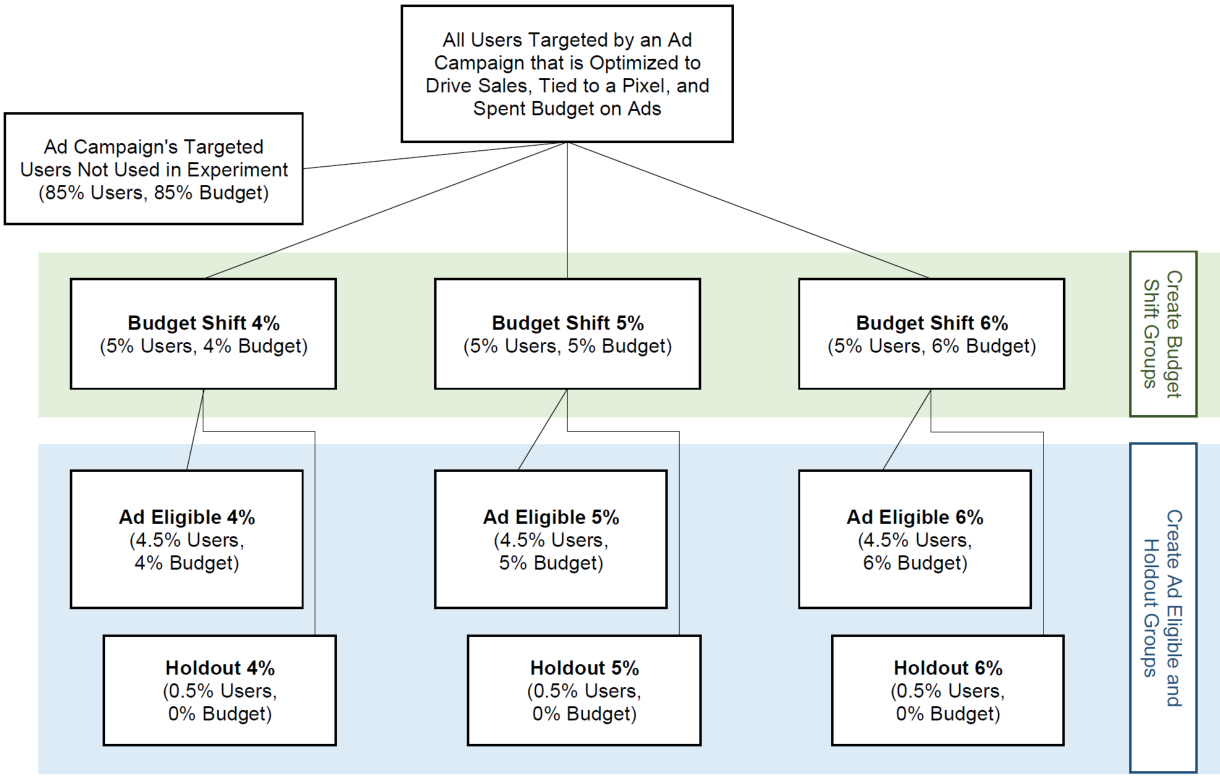
	(1) Historical Ad Spend	(2) Num. Ads	(3) Num. Campaigns	(4) Days Since First Ad
Panel A: Below Median				
Budget Spend	1.8213***	1.8424***	1.8900***	2.7239***
	(0.1462)	(0.2251)	(0.1968)	(0.2284)
N	270,238	272,408	261,352	526,576
R^2	0.575	0.592	0.606	0.705
Panel B: Above Median				
Budget Spend	3.3559***	3.4084***	3.4764***	3.6111***
	(0.1277)	(0.1320)	(0.1375)	(0.1536)
N	728,728	726,558	737,614	472,390
R^2	0.696	0.701	0.702	0.691

Table A7.4: Advertisers At Least 180 Days Old

	(1) Historical Ad Spend	(2) Num. Ads	(3) Num. Campaigns	(4) Days Since First Ad
Panel A: Below Median				
Budget Spend	1.8179***	2.1498***	1.9536***	2.5968***
	(0.1524)	(0.2753)	(0.1980)	(0.2154)
N	242,746	246,928	241,184	471,124
R^2	0.557	0.579	0.612	0.704
Panel B: Above Median				
Budget Spend	3.3587***	3.3922***	3.4577***	3.6722***
	(0.1297)	(0.1333)	(0.1381)	(0.1595)
N	663,488	659,306	665,050	435,110
R^2	0.700	0.702	0.703	0.693

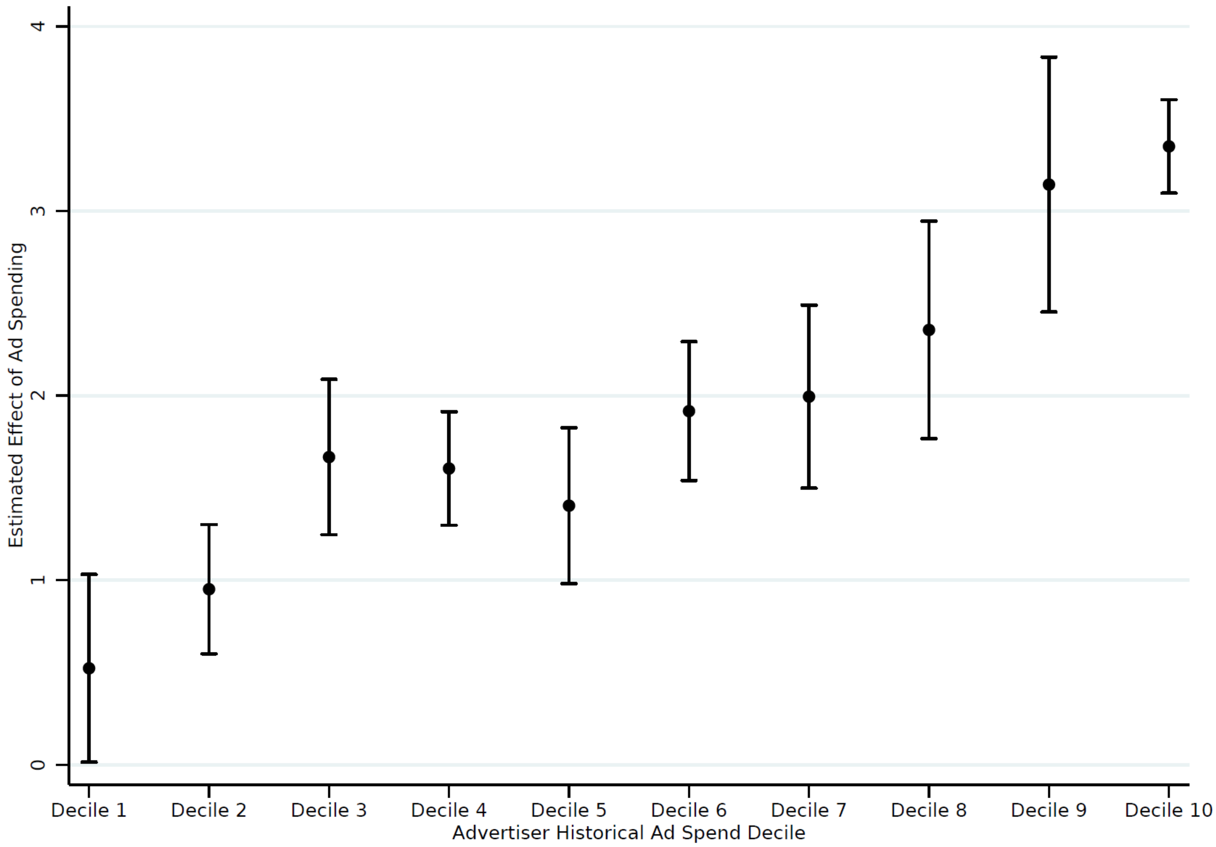
Notes: These tables report the results of estimating equation (1) for four different measures of advertiser experience, with the sample restricted to only those advertisers who created their first ad at least 30, 60, 90 and 180 days ago, respectively. N refers to the number of observations where each observation is recorded at the ad campaign – treatment group level. There were ad campaigns from 210,133 advertisers in the experiment. Column (1) reports results for our benchmark measure: the historical ad spending prior to the week of the experiment. Columns (2) - (4) report results for the number of ads, number of campaigns, and number of days since the advertiser first created an ad (a proxy for advertiser age). Panel A reports the results from regressions using the below median values of the experience variables of interest, while Panel B reports the same using above median values. The dependent variable in each panel is the revenue made by ad campaigns. Ad campaign fixed effects are included in each regression. Standard errors are in parenthesis. We cluster standard errors at the advertiser level. Statistical significance: * p<0.05, ** p<0.01, and *** p<0.001.

Figure A1: Experimental Design for Benchmark Results and Robustness Checks



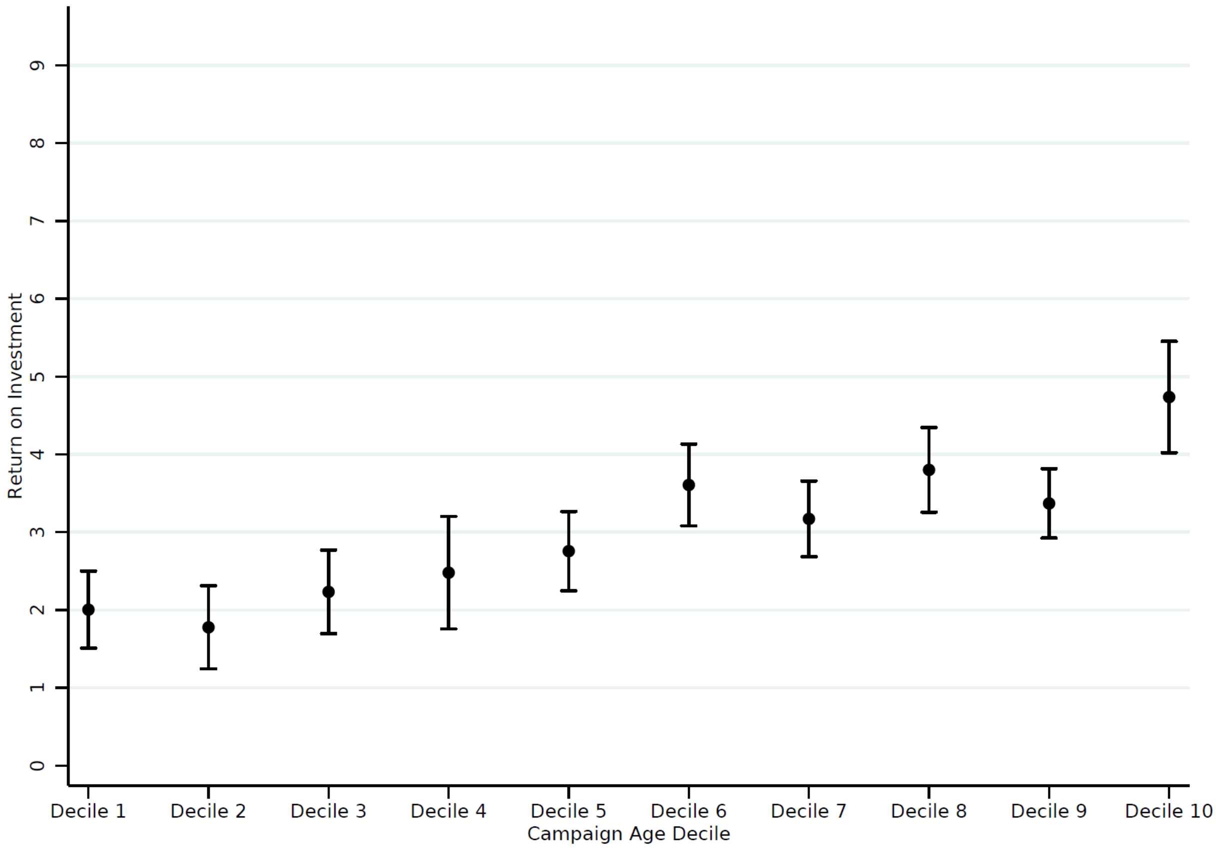
Notes: For each ad campaign in our experiment, 15% of its targeted users were randomly assigned to one of three approximately equal-sized Budget Shift groups; the remaining 85% of targeted users were excluded from the experiment for that campaign. Budget Shift groups were further subdivided into a group of Ad-Eligible users, who were given the opportunity to see ads from the campaign, and Holdout users, who were denied the ability to view ads from the campaign. The budgets allocated (internally) by the ad algorithms were approximately balanced between Ad-Eligible and Holdout users, which ensured that, within a Budget Shift group, Ad-Eligible and Holdout users were comparable. If an ad from an advertiser's campaign won an auction to appear in a user's feed, the ad was sent to the user's feed if that user was in an Ad-Eligible group and the minimum amount necessary to have won the auction was deducted from the advertiser's campaign budget. For Holdout users, the winning ad was replaced with the second-placed ad and the campaign budget that the first-placed advertiser would have spent on Holdout group users was allocated across users in the Ad-Eligible group (in the same Budget Shift group).

Figure A2: The Effect of Ad Spending by Historical Ad Spend Decile



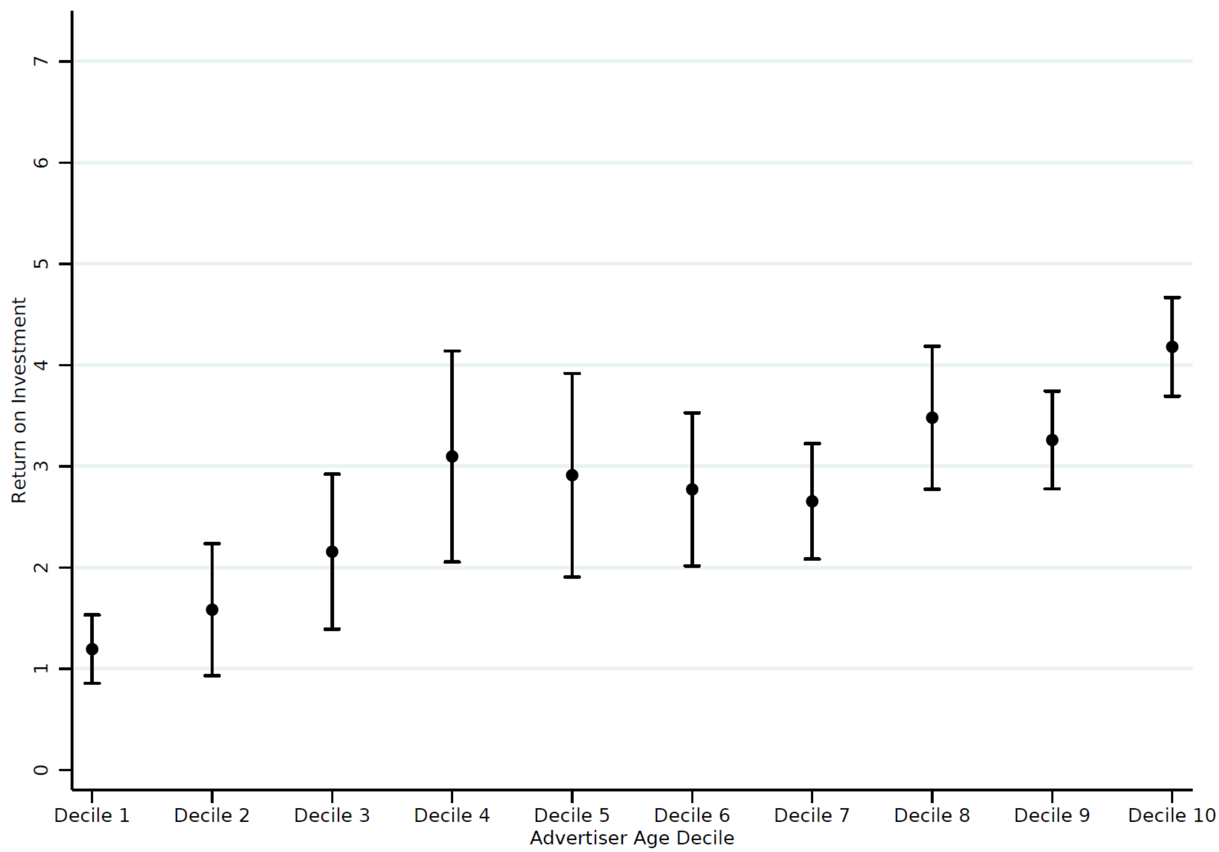
Notes: This figure presents the results of splitting the sample by decile of historical ad spending and estimating equation (1) for the Budget Shift 5% group observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure A3: The Effect of Ad Spending by Campaign Age Decile



Notes: This figure presents the results of splitting the sample by decile of campaign age and estimating equation (1) for the Budget Shift 5% group observations that fall within each quintile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

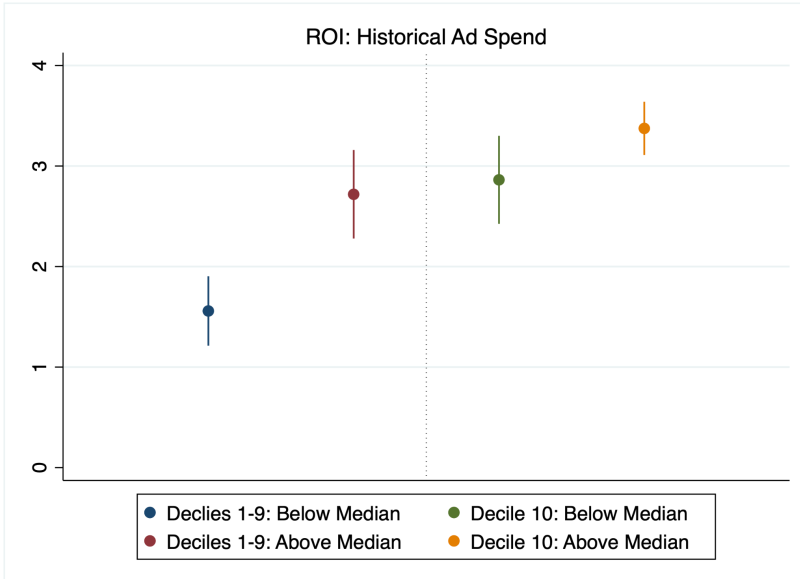
Figure A4: The Effect of Ad Spending by Advertiser Age Decile



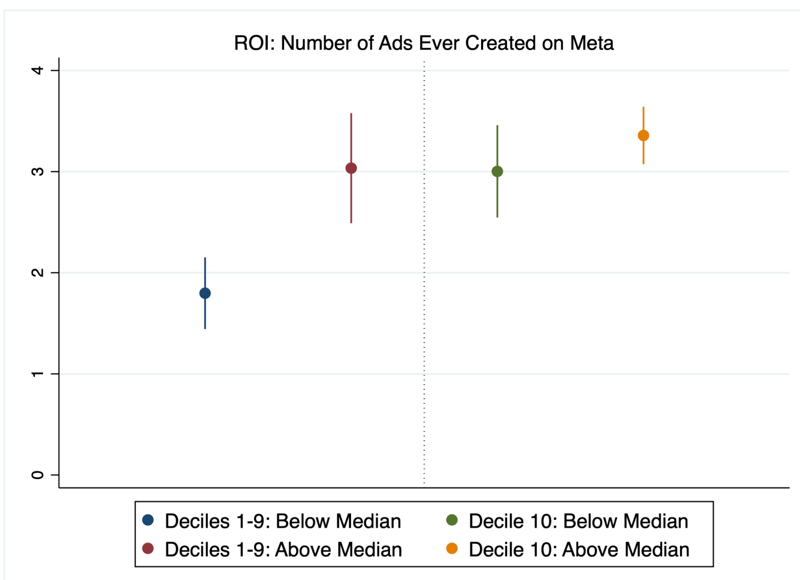
Notes: This figure presents the results of splitting the sample by deciles of advertiser age and estimating equation (1) for the Budget Shift 5% group observations that fall within each decile. Each bar shows the regression point estimate on Budget Spend and the corresponding 95% confidence interval.

Figure A5: The Effect of Ad Spending by Advertiser Experience: Separate Estimates for the Largest Spending (10th Decile) Advertisers

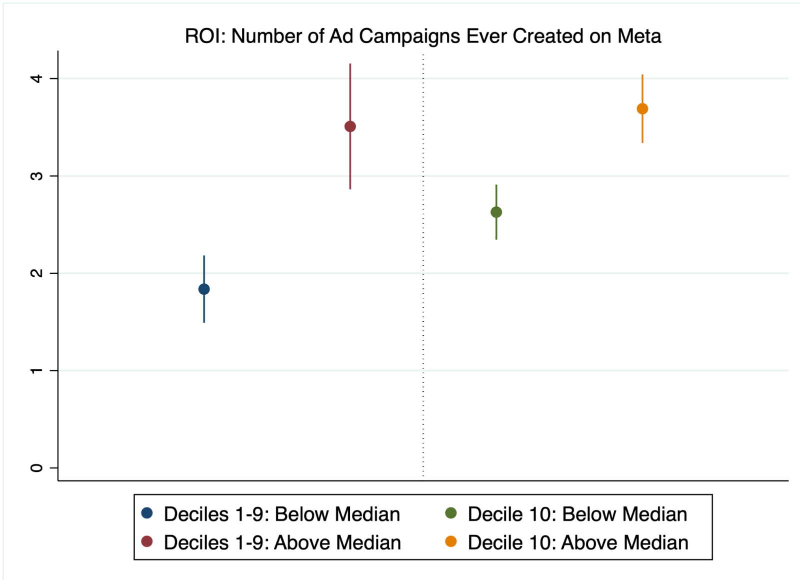
(a) Experience Measured by: Historical Ad Spend



(b) Experience Measured by: Number of Ads Created



(c) Experience Measured by: Number of Campaigns Run



(d) Experience Measured by: Days Since First Ad Created

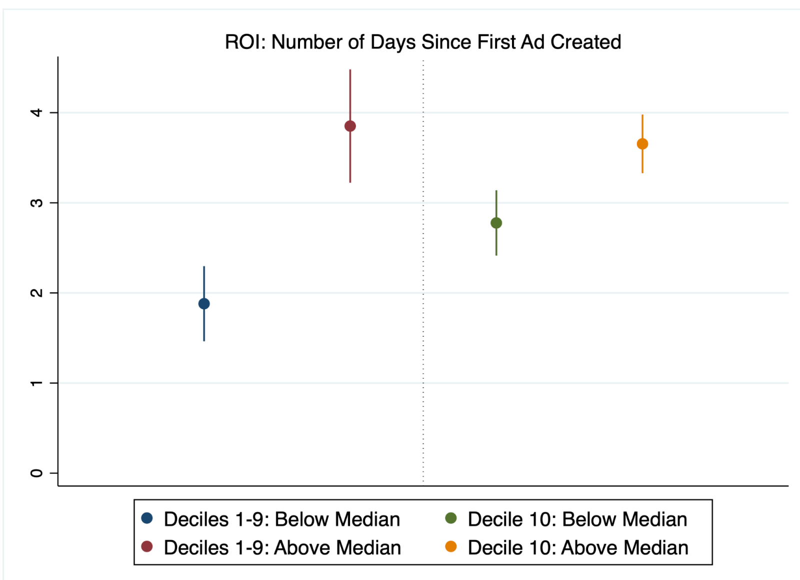


Figure A6: Average Ad Aesthetic Score Across Advertiser Age Buckets Over 0-12 Months: Non Parametric Estimation Using a Mann Kendall Test

Mann-Kendall Test parameters: tau = 0.515, 2-sided p-value = 0.0236