

Submitted to  
manuscript (Please, provide the manuscript number!)

# Worn-Out or Just Getting Started? The Impact of Frequency in Online Display Advertising

Randall A. Lewis

Google, Inc., randall@econinformatics.com,

Does repeated exposure to the same advertisement lead to wear-out, i.e., diminishing returns? This paper analyzes 2.8 billion impressions in 30 natural experiments on the Yahoo! Front Page to estimate the causal effects of ad frequency on users' clicks and conversions. The experiment randomly varies frequency, eliminating confounding selection bias from correlations between browsing behaviors and the propensity to click and convert. Failing to account for this selection bias can lead to erroneous estimates of wear-in, for which no evidence is found, or show excessive wear-out for campaigns with little to no wear-out.

Wear-out is heterogeneous for clicks: four out of 30 campaigns exhibit significant wear-out after one or two exposures while ten campaigns show little after as many as fifty. Those campaigns that experience the most wear-out also have the highest click-through rates, suggesting that wear-out may signal fresh creatives and renewed interest by the exposed audience. A model for new-account sign-ups finds comparably modest wear-out for both clicks and new account sign-ups for one advertiser's campaigns. Finally, a discussion highlights that wear-out, whether absent or present, is a crucial part of quantifying an advertiser's marginal return from advertising.

*Key words:* Online display advertisement, frequency, reach, identification, advertising effectiveness, natural experiment, false synergy, wear-out

---

## 1. Introduction

Online advertising runs on repetition. Tens of billions of dollars are spent each year on repetitive online media such as email, search, display, audio, and video ads intended to reach consumers with information and enticements. In the United States during 2013, \$42.8 billion were spent advertising online with display-related ads accounting for 30% or \$12.8 billion.<sup>1</sup> Despite the scale of these advertising expenditures and the voluminous repetition of identical messages to consumers, little is known about the marginal effectiveness of these messages. Advertisers need to understand how exposure frequency influences consumer behavior. Frequency forms the primary determinant of the cost of advertising: how many times can I send the message and still obtain a profitable return? As the number of messages increases, the advertiser can experience increasing, decreasing, or constant

<sup>1</sup> IAB Internet Advertising Revenue Report 2013: <http://iab.net/AdRevenueReport>.

returns to scale. Increasing returns to scale is commonly referred to as “wear-in” or “synergy” while decreasing returns is known as “wear-out” Pechmann and Stewart (1988).

The main challenge facing a study of ad frequency is that advertisers intentionally target their ads to responsive subpopulations. As a result, the people that are thought to be more likely to be influenced by the ad are more likely to see a large number of impressions. This confounds the analysis of frequency and tends to bias estimates toward synergy. Past studies have attempted to overcome this problem by using a controlled laboratory setting or trying to account for selection using econometric procedures. While this research has furthered our understanding of advertising, both approaches have well-known drawbacks. Laboratory studies are criticized on external validity. Observational studies have a difficult time effectively accounting for selection bias (Lewis et al. 2011). In this paper, I use a natural experiment described in section 2 that induced exogenous variation in ad frequency while retaining the natural decision-making environment of the field. Visitors to the Yahoo! Front Page are shown a different ad depending on whether their webpage loaded on an even or odd second. This pseudo-random ad exposure, when coupled with data on clicks and sign-ups for the same users, allows me to identify the causal effects of frequency on these outcomes for a number of online display ad campaigns.

The Yahoo! Front Page is a major internet portal where 30–40 million users visit each day, amassing between 140 and 200 million total page views. For each page view the primary ad unit is refreshed, delivering the user one ad impression. The odd- and even-second rotation between two advertisers on certain days creates exogenous variation in the number of ads each user sees for both advertisers. However, the total number of ads seen is not exogenous—it is determined by the number of times the user decides to visit [www.yahoo.com](http://www.yahoo.com). If the number of visits to the Yahoo! Front Page is correlated with unobserved heterogeneity that is correlated with ad responses, then models which fail to account for this endogeneity will produce biased estimates of the impact of frequency.

I have anonymous individual-level data on ad views and ad clicks for 30 campaigns shown on 15 days from January through March 2010 on the Yahoo! Front Page. I present a model in section 3 to estimate the effects of frequency on clicking behavior. The model accounts for the unobserved heterogeneity in number of visits when estimating the effects of frequency on a user’s likelihood to click on the ad to follow the link to the advertiser’s website. As a result, the model only exploits the exogenous variation in the number of ads seen by users who visit the page a given number of times.

While clicks are commonly analyzed by researchers, conversions are more interesting to study because they represent economic activity which directly affects the advertiser’s bottom line. To accommodate conversions, I adjust the restrictions on the model to estimate the effects of that

campaign's ad frequency on an individual's likelihood of signing up for a new account. The online environment allows me to gather data on both the number of impressions seen and whether the user signed up for an account on the advertiser's website.<sup>2</sup> I combine four of the advertiser's campaigns to examine the effects of frequency on conversions.

I find in section 5 that there is widely-varying heterogeneity in the effects of frequency on clicking behavior. Four of the 30 campaigns wear out exceedingly fast, while ten exhibit near constant returns to scale for more than 20 impressions. For example, several campaigns achieve roughly 40 times the impact of showing one ad from showing 40 ads while other campaigns achieve only 3 times the impact of showing one ad from showing 40. I show that naively using observational data and ignoring selection bias can lead to large biases in the frequency estimates; for clicking on the Yahoo! Front Page, the bias overstates wear-out for 26 of the 30 campaigns. In addition to sizable over- and under-statements of the estimates of wear-out, erroneous findings of synergy (increasing returns to scale) are also common—half of the campaigns find synergy from showing two ads when compared to showing just one. Using exogenous variation in ad delivery and accounting for unobserved heterogeneity, I find no evidence of synergy. The analysis of frequency on conversions shows a positive return to frequency for as many as 20 impressions.

## 2. Related Literature

The effects of ad frequency have been studied with both field and lab experiments. An early influential field experiment by Zielske (1959) measured the impact of direct mail ad frequency on housewives' recall of a low-market-share brand of flour over the course of a year. Each woman was mailed thirteen ads, each spaced by either one week or four weeks. Brand-name recall decayed exponentially over time, and four-week intervals yielded higher average brand-awareness over the course of the year than did the one-week intervals—a result consistent with decreasing returns to frequency on recall Simon (1979). Lab experiments have also been used to study the effects of repetition of print media on recall (Craig et al. 1976) and of television and internet ads on attitudinal measures (Campbell and Keller 2003). Their findings corroborate Zielske's field experiment in finding decreasing returns to frequency—and potentially even negative marginal effects of frequency from high levels of exposure.

In a pioneering observational field study, Tellis (1988) combined data on television ad exposure with scanner data for consumer purchases to evaluate the impact of advertising frequency on sales. Using a Tobit analysis to correct for selection biases, he concludes that the optimal frequency for television is perhaps two or three commercials per week. However, he cautions that extending the

<sup>2</sup> "Conversion pixels" are 1-pixel images on websites commonly used for web analytics. Here, the 1-pixel image was downloaded when the user completed the online sign-up process.

results beyond those frequencies is difficult because the study lacks sufficient data to make any claims beyond four or five exposures.

With the widespread adoption of the internet in the mid-1990s and the proliferation of online display and search advertising, many researchers harnessed new technologies to more effectively explore the impact of ad frequency on behaviors. Early observational research by Chatterjee et al. (2003) studied click-through rates (CTRs), or the fraction of ads downloaded that were clicked on by users in order to visit the advertiser's website. They found wear-out in that increasing frequency led to a decline in CTRs. Kameya and Zmija (2002) outlined other early research using observational data which found that after five impressions, returns are diminishing (Carlson and Hislop 2001, Morgan Stanley Dean Witter 2001). Dreze and Hussherr (2003) used surveys and eye-tracking equipment to study online display advertising in 1999. They found significant "ad blindness" among participants, where users had learned to avoid looking at certain areas of the webpage because they knew what looked like an ad versus the content that they were seeking. In particular, they found that the observed 50% likelihood of users looking at the ad influenced consumers' brand awareness and recall. Havlena et al. (2007) combined online display advertising frequency measurements with TV and print advertisement frequencies and found positive returns to frequency on several survey-based measurements such as recall.

The combination of internet technologies and field experiments has highlighted the endogeneity of advertising frequency and led to important findings. Using a large-scale field experiment, Lewis and Reiley (2014) found that online display advertising targeting a nationwide retailer's customers produced a statistically and economically significant positive effect on online and offline sales. Yet, those who visited Yahoo! and saw the retailer's ads actually purchased less, on average, than those who were eligible to see the ads but failed to browse online enough to see any. Not surprisingly, then, a naive regression of each customer's sales on the number of ads seen produced a negative slope, falsely suggesting that advertising stifles sales. This analysis, similar to many field studies, incorrectly assumes that the number of ads seen is exogenous even though users who visit Yahoo! more frequently are shown more ads. In two more recent experiments, Johnson et al. (2014) measure the causal impact of frequency on both online and offline sales by building upon this paper, and Sahni (2013) measures the effects of frequency on a restaurant search website with a similar design as in this paper.

In summary, advertising frequency has been studied in a variety of media, but several shortcomings of lab experimentation and observational studies have limited the solidity of the conclusions. Digital field experiments facilitate clean measurements of the effects of frequency. By identifying exogenous variation in the number of ads shown to each person and observing their behavioral responses, I am able to rigorously quantify the effects of frequency.

### 3. Identification Strategy

The Yahoo! Front Page for the United States market, shown in Figure 1, is a major internet portal for 40 million unique visitors each day, providing up-to-date news and content which appeals to most demographics. The ads on the Yahoo! Front Page provide a unique opportunity to study frequency due to their single display location, their stable viewing population, their lack of competing ad units, and most importantly, their unique sales strategy which creates a natural experiment.

Yahoo! Front Page advertisements are sold either as a “roadblock,” where every visit to the page on a specific date includes the ad from one exclusive advertiser, or as an “ad-split,” where half of all visits to the page includes the advertiser’s ad. The two halves are split by whether a visit happens on an odd second (e.g., 9:00:01) or an even second (e.g., 9:00:02). By purchasing an ad-split, an advertiser pays for only half of that day’s display impressions but reaches 77% of visitors with at least one ad—a balance between reach and frequency. Importantly, the Yahoo! Front Page ad server only uses this very simple odd-even-second targeting rule—ignoring the identity of the user when deciding which ad to serve. As long as there are no systematic differences between odd- and even-second visits, ad delivery to each user on “split” days varies exogenously, providing variation in ad frequency between the two advertisers in the ad-split.

On an ad-split day, individuals who visit the Front Page ten times see between zero and ten impressions from the “odd-second” advertiser and the complement to ten from the “even-second” advertiser. The top panel of Figure 2 illustrates this binomial distribution empirically for one advertiser and shows that comparing groups of users who randomly saw different numbers of an advertiser’s reveals the causal impact of ad frequency. Focusing on this random variation is important because unobserved heterogeneity across users of differing browsing intensities is correlated with the number of ads seen, confounding ad effect estimates. Simply stated and illustrated in the bottom panel of Figure 2, people who see more ads (e.g., because they use the internet more) differ in how they respond to the ads relative to those who see fewer. In this case, individuals who frequent the Yahoo! Front Page are less prone to click on an ad or sign up for an advertised online service than those who only visit occasionally. By tailoring a model to this natural experiment, I can hold the number of visits to the page fixed and observe responsiveness to the ads under different numbers of exposures.

### 4. Model

I present a simple semiparametric model for the effects of frequency on behavior. I first consider restrictions best suited for modeling ad clicks and then adapt those restrictions to accommodate additional outcomes, such as new account sign-ups.

#### 4.1. Click Model

Let  $\theta$  denote an individual’s browsing “type.” Here, this browsing type,  $\theta$ , is defined as the total number of opportunities to be shown either ad-split advertiser’s ad and is assumed to be exogenous. I model the expected number of clicks  $C$  by individual  $i$  of type  $\theta$  when shown  $f$  impressions from advertising campaign  $c$  as

$$E[C_{ic}|f, \theta] = \bar{b}_c \cdot b_c(\theta_{ic}) \cdot h_c(f_{ic}) \quad (1)$$

$$= e^{\bar{a}_c + a_c(\theta) + g_c(f)}. \quad (2)$$

This model includes several parameters of interest. In equation 1,  $\bar{b}_c$  is the  $c^{\text{th}}$  campaign’s scaling constant—analogue to the campaign’s baseline click-through rate. The function  $b_c(\cdot)$  defines the average heterogeneity in clicking propensity across individuals of different browsing types for the  $c^{\text{th}}$  campaign. I call this the “relative heterogeneity” function. The function  $h_c(\cdot)$  defines how the expected number of clicks varies with the number of ads delivered and is dubbed the “relative frequency” function. These parameters can be estimated via the Poisson regression model in equation 2 and exponentiating the point estimates and computing standard errors via the delta method or bootstrap simulations.

Poisson regression requires standard exogeneity assumptions to make use of the ad-split natural experiments. Specifically, users’ clicking behavior is only correlated with  $\theta_{ic}$  and  $f_{ic}$  through the functions  $b_c(\cdot)$  and  $h_c(\cdot)$ . This assumption is at the core of the exogenous variation induced by the ad-split ad delivery that, for a given  $\theta_{ic}$ , produces a range of values for  $f_{ic}$  which identifies the causal relative effects of frequency function,  $h_c(\cdot)$ , separately from  $b_c(\cdot)$ .

The frequency response function,  $h_c(f)$ , explains how changing the number of ad exposures influences the relative willingness of an individual to click. Individual heterogeneity across browsing types scales this function up or down, depending on  $b_c(\theta)$ . This multiplicatively separable structure is appropriate given the shapes of the fully nonparametric frequency response curves seen in the bottom panel of Figure 2 which appear to scale up in  $\theta$  for that campaign. However, as a result, both  $b_c(\cdot)$  and  $h_c(\cdot)$  are only separately identified up to scale. I normalize  $b_c(\theta) = 1$  and  $h_c(f) = 1$  at  $\theta = 1$  and  $f = 1$ —the modes of the  $\theta$  and  $f$  distributions. These restrictions imply that the scaling factor  $\bar{b}_c$  is equal to the click-through rate for users of type  $\theta = 1$  who have seen  $f = 1$  ads for the  $c^{\text{th}}$  campaign. These restrictions are trivially imposed by simply using  $\theta = 1$  and  $f = 1$  as the base categories in the Poisson regression.

I estimate the Poisson regressions with nonparametric specifications of  $b_c(\cdot)$  and  $h_c(\cdot)$  with  $1 \leq \theta \leq 120$  and  $0 \leq f \leq 70$ . I use a combination of dummy variables and linear basis splines to provide full flexibility of the estimates for small values of the functions’ arguments and sufficient flexibility

and precision for larger values. For  $b_c(\cdot)$ , dummy variables are included for  $\theta = 1$  to 15, and b-splines span values of  $\theta$  from 15 to 120. Regarding  $h_c(\cdot)$ , I include dummies ranging from  $f = 0$  to 15 and b-splines from 15 to 70.<sup>3</sup> All standard errors are heteroskedasticity-robust to prevent any Poisson distributional assumptions from affecting inference of the conditional mean estimated by Poisson regression.

## 4.2. Conversion Model Restrictions

In the click model,  $h_c(0) = 0$  because a user cannot click without first viewing the ad. By allowing  $h(0)$  to differ from zero, the model can also measure the effects of advertising frequency on other outcomes of interest to advertisers such as branded search queries, page views or transactions on the advertiser's website, or online and offline sales. The more general conversion model for an outcome,  $Y_{ic}$ , is

$$E[Y_{ic}|f, \theta] = \bar{b}_c \cdot b(\theta_{ic}) \cdot h(f_{ic}). \quad (3)$$

This model is equivalent to the click model, except  $h_c(0) = 1$  is now the location and scale restriction. All relative frequency effects are no longer in terms of the impact of one impression ( $h_c(1) = 1$  was the click model restriction), but rather in terms of the baseline sign-up rate. Additionally, to more precisely measure the relative frequency effects for a single advertiser, I restrict both  $b_c(\cdot)$  and  $h_c(\cdot)$  be equal across that advertiser's similar campaigns. However, to avoid bias from cross-campaign variation in outcome levels, I do not restrict the baseline sign-up rates,  $\bar{b}_c$ .

I use this model to examine the effects of frequency on new account sign-ups in section 6.4. As with the click model, I combine dummy variables and linear b-splines to estimate the model using Poisson regression on visitors with ad type  $1 \leq \theta \leq 100$  and between 0 and 40 impressions. For  $b(\cdot)$ , dummy variables are included for  $\theta=1$  to 4, and b-splines span values of  $\theta$  greater than 4 and up to 100. Regarding  $h(\cdot)$ , I include dummies ranging from  $f = 0$  to 2 and b-splines from 2 to 40.<sup>4</sup> Again, the first-order splines simplify estimation while providing sufficient flexibility.

## 4.3. Model Discussion

A naive model might compare users who see different numbers of ads without accounting for the intrinsic differences across those users, potentially introducing endogeneity. The purpose of this model is to accommodate the exogenous variation identified in the natural experiment and both quantify and guard against such endogeneity in the frequency estimates. The two primary requirements of the model are the introduction of  $b(\theta)$  and the conditional independence of the residual, given  $\theta$  and  $f$ .

<sup>3</sup> For  $b(\cdot)$ , the b-splines include a constant for  $\theta > 15$  and then knots at 15, 20, 25, 30, 40, 50, 60, and 90. For  $h(\cdot)$ , the b-splines include a constant for  $f > 15$  and then knots at 15, 20, 25, 30, 40, and 60.

<sup>4</sup> For  $b(\cdot)$ , the b-splines include a constant for  $\theta > 4$  and knots at 4, 7, 14, 19, 25, 30, 35, 40, 45, 50, and 60. For  $h(\cdot)$ , the b-splines include a constant for  $f > 2$  and knots at 2, 4, 7, 10, 14, 19, and 25.

The  $b(\cdot)$  function in equation 1 allows users of different browsing intensities to have different absolute levels of the outcomes, thus respecting their heterogeneity. However, a naive click model such as

$$E[C_{ic}|f, \theta_{ic}] = \bar{b}_c \cdot h_c(f_{ic}) \quad (4)$$

$$= e^{\bar{a}_c + g(f_{ic})} \quad (5)$$

that sets  $b(\theta) = 1$  for all browsing types needs the number of ads sees to be exogenous for standard estimation methods to be valid. However, if visitors of various types differ in other ways that affect their propensity to click, regression estimates of  $h(\cdot)$  will be biased.

The conditional independence of the residual can fail for a variety of reasons. For example, if visitor  $i$  shows up and clicks the first ad that they see for the advertiser and go to that website, never to return to the Yahoo! Front Page; however, visitor  $i$  would have seen ten ads had she not seen that advertiser’s ad. Hence, the model may be misspecified because it ignores that an individual’s click may distract them from returning, so  $\theta_{ic}$  may be mismeasured and  $f_{ic}$  may not be exogenous.<sup>5</sup> In short,  $\theta_{ic}$  and  $f_{ic}$  could be endogenous to whether a user clicks. Future research should investigate the empirical relevance of these endogeneity concerns.

One behavioral model consistent with the independence assumption is that user  $i$  decides to visit the Yahoo! Front Page  $\theta_{ic}$  times in a given day at exogenously set times and squeezes in around that preset schedule any additional browsing that results from clicking on the ad. While this likely is not a perfect description of true behavior, it probably is a good approximation. However, all frequency analyses on field data depend on user behavior for ad delivery and are subject to this criticism as the response to the most recently delivered ad could influence whether the next ad is delivered at all. In spite of these challenges, the combination of the natural experiment and modeling assumptions to account for heterogeneity across visitors is a significant step toward accurately measuring the effects of advertising frequency on behaviors.

## 5. Data

The Yahoo! Front Page ad-split data includes individual-level impression and click data for 30 campaigns and one advertiser’s online account sign-up data during that advertiser’s four campaigns. The individual-level impression and click data was obtained for fifteen days during January, February, and March 2010 when two advertisers participated in ad-splits, providing a total of 30 campaigns to study. I observe each user’s anonymous unique identifier along with the number of

<sup>5</sup> The severity of such effects are likely second order to the first-order endogeneity in the number of exposures. Future research could examine “click-stream” data to quantify the severity of this bias: if the likelihood of returning to the Yahoo! Front Page depends upon which ad you were shown, then this would suggest that the browsing type,  $\theta$ , may be mismeasured.

ad views,  $f$ , and ad clicks for both ad-split advertisers. This allows me to construct each user's browsing type,  $\theta$ , which denotes, "How many ads could this user have seen today for each advertiser?" With this information, I can compare outcomes such as clicks for users of the same type,  $\theta$ , who were randomly shown different numbers of impressions,  $f$ .

The online new account sign-up data comes from a conversion pixel service which Yahoo! provides to advertisers. The one advertiser collected this information during four campaigns during the first half of 2010, one of which is part of the 30 campaigns in the individual-level impression and click data. The data indicates which unique user signed up for a new account on either the day of the campaign or the following day. This data provides a unique opportunity to learn how the effects of frequency extend beyond just clicks to conversions.

Table 1 shows summary statistics for the 30 campaigns in the primary analysis. On each of the 15 days, roughly 40 million unique visitors trafficked the Yahoo! Front Page, viewing an average of 4.6 ads per day, split between the two advertisers. Each campaign delivered between 70 and 98 million impressions and attracted roughly 100,000 "clickers"—unique users who clicked on the advertiser's ad. Analyzing clickers to eliminate duplicate clicks will mechanically shift estimates toward wear-out because users who browse longer will only have their first clicks counted. However, because the difference between the total numbers of clicks and clickers is less than 10% for each campaign in Table 1, multiple-clickers form a small share of the clickers and do not materially affect the findings.

There is substantial heterogeneity in the response rates to different advertisers: some advertisers were able to attract clicks from as many as 1.36% of all visitors to the Yahoo! Front Page, while most others attracted between 0.10% and 0.30%. The campaign numbers, which appear out of order, correspond to the click-through rate (CTR) ranking from highest to lowest and is used for organizing graphs of the main findings. In total, I observe 1.2 billion visitor $\times$ campaigns which equate to 609 million visitor $\times$ days—two campaigns per day. These visitor $\times$ days included only 237 million unique visitors, or roughly 39% of 609 million. Of the unique visitors, 146 million visited only one of the 15 days, 7 million visited five days, 2 million visited ten days, and 2 million showed up all 15.

The distribution of browsing types,  $\theta$ , is highly-skewed. For example, on day 5, 90% of visitors saw fewer than 10 impressions for both ad-split advertisers. While visitors saw an average of 4.7 ads that day, the skewed distribution places the median on the margin of 2 and 3 impressions. Under usual circumstances such skewed data would leave little hope of identifying any effects out as far as 20 or 30 impressions; however, this data produces a remarkable 5,780 unique users who were exposed exactly 50 times ( $\theta = 50$ ) on the Yahoo! Front Page that day. Hence, the scale of the data facilitates reasonably tight confidence intervals for model estimates for ad frequencies as large as  $f = 40$ .

## 6. Analysis of Clicks

I estimate the cumulative effects of display ad exposure on ad clicks to measure the degree of increasing, constant, or decreasing returns to scale for the 30 ad campaigns shown to users on the Yahoo! Front Page. I use the click model described in section 4, repeated here:

$$E[C_{ic}|f, \theta] = \bar{b}_c \cdot b_c(\theta_{ic}) \cdot h_c(f_{ic}). \quad (6)$$

Most campaigns represent distinct advertisers, and several days separated any campaigns from the same advertiser. However, each Yahoo! Front Page campaign does not necessarily account for all ads being shown by the advertiser—I estimate the frequency effects of showing these ads on this online location, taking all other contemporaneous advertising for this advertiser as given.

### 6.1. Effects of Frequency on Clicks

Figure 3 shows the relative frequency effect,  $h(f)$ , the cumulative performance from  $f$  ads for all 30 campaigns with 95% confidence intervals. The  $h(1) = 1$  scale normalization facilitates scale-free comparisons of wear-out across campaigns. Two features immediately stand out. First, several campaigns experience rapid wear-out, with one campaign’s estimate showing  $h(5) = 1.7$ , or roughly 60% of the impact of showing five ads is from the first. These ads are seen and clicked within a small number of impressions, experiencing rapidly decreasing returns to scale. Second, many campaigns experience relatively little wear-out, with  $h(20) = 20$ , implying that twenty times the impact of showing one ad is achieved by showing twenty. For these advertisers, each incremental impression is just as potent as the last; the ads experience constant returns to scale. Further, recalling that the campaigns are numbered in descending order by CTR, Figure 3 shows an interesting relationship between absolute campaign performance and wear-out: higher CTR campaigns (e.g., campaigns 1, 2, 3, ...) generally exhibit clicker rates which taper off much more rapidly with additional impressions than the lower CTR campaigns.

Among the 30 campaigns, there are ten campaigns with constant returns to scale ( $h(20)$  in  $[18,21]$ ), six with mild wear-out ( $h(20)$  in  $[15,18]$ ), ten with moderate wear-out ( $h(20)$  in  $[12,15]$ ), and four with extreme wear-out ( $h(20)$  in  $[8,12]$ ). In words, the benefit of showing twenty impressions ranges from 1.7 to 4.1 times the impact of showing five impressions—from strongly decreasing returns to scale to constant returns to scale. The four campaigns in Figure 5 illustrate the varying degrees of relative frequency performance for the 30 campaigns. The top left panel provides an example of a campaign where the estimate of the relative frequency curve follows the  $Y = X$  line in the plot which coincides with constant returns to scale. The top right panel illustrates the mild wear-out as evidenced by the increasingly large deviation from the  $Y = X$  line for higher frequencies. The bottom left panel shows moderate wear-out, with an even larger deviation from

constant returns. However, the bottom right panel shows a campaign where the wear-out is so extreme that an additional 35 ads are necessary to equal the impact of the first five.

None of the curves suggest negative impacts from showing too many impressions to users within a single day (commonly known as the “inverted U”) because all marginal effects of ads on clicking are positive. However, if marginal willingness to click did turn negative, in a “hold-up” fashion, the advertiser may have already received the clicks that were expected from showing a smaller number of impressions earlier. Thus, heterogeneity in attitudinal effects could be both positive and negative as interest in clicking dwindles for some viewers, but the frequency estimates for clicks would be unable to illustrate the negative impact of overexposure to the ad. While the level of statistical precision does not rule out constant effects for relevant frequencies of several of the campaigns, additional attitudinal or direct-response (e.g., sales and sign-ups) outcomes should be used to investigate the risks of overexposure to the ads in future work.

## 6.2. Browsing Type Heterogeneity

This natural experiment allows us to see how browsing heterogeneity affects wear-out estimates. Importantly, I find that the relationship between browsing types and users’ propensity to click varies widely from campaign to campaign. The browsing type heterogeneity function,  $b(\theta)$ , identifies differences in light and heavy users’ willingness to click on ads for each campaign. Figure 4 shows the estimated functions for each of the 30 campaigns analyzed, where each has been normalized such that  $b(1) = 1$  (i.e., relative to users who only visit the Yahoo! Front Page once during the day). In the absence of heterogeneity, the number of visits to the Yahoo! Front Page would be uncorrelated with clicking and the entire function should be constant and equal to one. However, most of the campaigns exhibit substantial variation in the relative propensities to click with some campaigns showing decreasing propensities to click in browsing intensity and others showing increasing click propensities.

I examine four campaigns whose curves are representative of the range of the browsing type heterogeneity estimates which I have divided into four<sup>6</sup> categories presented in Figure 6. The 30 campaigns are divided into fifteen for which light users respond more (top left panel;  $b(50)$  in  $[0.45, 0.95]$ ), eight for which responsiveness is roughly constant across browsing types (top right panel;  $b(50)$  in  $[0.95, 1.25]$ ), five for which heavy users respond more (bottom left panel;  $h(50)$  in  $[1.25, 1.75]$ ), and two for which heavy users respond much more than light users (bottom right panel;  $h(50)$  in  $[1.75, 2.25]$ ). Several campaigns experience one-half the responsiveness between light and heavy users as illustrated in the top left panel where users with browsing types of 100 are

<sup>6</sup> These four categories only represent a simple comparison between users of browsing type 1 and 50. Several campaigns exhibit more complex shapes such as curves rising sharply for the users of types one, two, or three but then slowly declining in users’ browsing type.

half as likely to click on the ad, holding the relative frequency effects and CTR constant. Others experience near-constant responsiveness for all users ranging from light to heavy users as shown in the top right panel of Figure 6—for these campaigns, browsing-type heterogeneity appears to not matter much. Finally, as illustrated in the bottom two panels, some campaigns experience greater response from heavier users, with some campaigns experiencing as much as 2.5 times the clicking responses from heavy users relative to the light users ( $\theta = 100$  versus  $\theta = 1$ ). Nevertheless, 90% of visitors to the Yahoo! Front Page are of browsing type 10 or less, and a closer look at this important portion of the browsing type curves in Figure 4 reveals that the curve is greater than one for virtually all campaigns. Therefore, users who visit between two and ten times are, on average, more willing to click than users who visit only once, controlling for frequency effects.

Ignoring this widely-varying, cross-campaign heterogeneity in clicking propensities across users of different browsing intensities leaves biased frequency estimates. Further, no simple correction, such as multiplying by a constant, can account for user heterogeneity when estimating the impact of frequency on clicking. The browsing type heterogeneity function is, perhaps, the simplest way to reliably account for user heterogeneity at the campaign or advertiser level.

The browsing type heterogeneity function is a composition of two factors: differences in the level of intrinsic interest in the ad’s offering and in the advertiser’s stock (Palda 1965, Mann 1975). Interest in the advertiser’s offering may be correlated with browsing intensity. For example, advertisements for online games might appeal to heavier internet users while ads for consumer packaged goods might appeal to lighter internet users who enjoy grocery shopping. That said, if an advertiser frequently advertises online, heavier users will see more ads and, hence, have a larger ad stock. This could cause the marginal responsiveness of heavier users to be much lower, especially if they have already responded to the company’s offering. Consequently, their willingness to click may be much lower—not because they were not good candidates for the ads, but because they had already responded. Regardless, the shape of  $b(\theta)$  can signal whether an advertiser’s budget is better spent on light or heavy users of the Yahoo! Front Page because the relative differences in responsiveness translate into absolute differences, after controlling for differences in expected wear-out.

### 6.3. Estimation Ignoring Heterogeneity

After accounting for browsing heterogeneity, it is possible to impose the constraint of  $b(\theta) = 1$  for all  $\theta$  and obtain the naive estimates. Imposing this constraint is equivalent to asking the observational question: do visitors who see different numbers of ads click differently? This question directly contrasts with the causal frequency question of interest: how does the same visitor’s behavior vary

when shown different numbers of ads? I compare the estimates obtained accounting for heterogeneity with naive estimates which assume that the number of ads seen is exogenous and, hence, that  $b(\theta) = 1$  for all  $\theta$ . The estimates for all 30 campaigns are plotted in Figure 3.

Four of the campaigns, shown in Figure 7, illustrate the range of conclusions resulting from naively assuming that all internet users are the same. In the top left panel, one campaign observed to be experiencing constant returns to scale when correctly accounting for the heterogeneity shows significantly decreasing returns to scale when ignoring the differences in browsing types. Many other campaigns show this effect, but to a lesser extent as in the top right panel. Still, several campaigns would have realized nearly the same results using either method—this is the case for the bottom left panel—while other campaigns actually wore-out faster than the naive estimates would suggest. This means that, by ignoring browsing type differences, campaigns that are experiencing constant returns to scale could be incorrectly designated as experiencing decreasing return to scale while other campaigns that are wearing out fast could be miscategorized as experiencing wear-out more slowly. Both of these mistakes are potentially costly.

For the 30 campaigns analyzed, biases resulting in the overstatement of wear-out are severe. Using the two estimates, I compute bias ratios of  $h(40)$  for the 30 campaigns. Only four campaigns' naive estimates understated the wear-out. The remaining 26 campaigns' wear-out was overstated by the naive estimator with 16 campaigns experiencing biases ranging between -32% and -62%. This significant overstatement of wear-out made by the naive model highlights the importance of accounting for user heterogeneity when estimating the effects of online advertising.

Finally, I examine an assertion frequently heard in industry Chang and Thorson (2004): there is a convex portion of the frequency response function,  $h(f)$ , where there are increasing returns to scale, or ad “synergy” from showing multiple ads because exposure to a few impressions is necessary for visitors to notice the ad’s message and overcome their reluctance to click. I examine the frequency response functions more closely to assess this claim by comparing the relative frequency effects of showing one ad versus two for the 30 campaigns. Figure 8 shows a simple comparison of  $h(2) - 2 * h(1)$  for the unconstrained and constrained (naive) estimators asking, “Is the likelihood of clicking on the second ad more than twice the likelihood of clicking on the first?” The unconstrained estimator only finds two campaigns (in the first quadrant) that might exhibit weak, but statistically insignificant synergy from showing the second ad, with the largest of the two representing a 1.0% increase in  $h(2)$  from synergy. However, the naive estimator designates half of the 30 campaigns (first and second quadrants) as exhibiting synergy with 3–6% boosts in responsiveness to the second ad,  $h(2)$ , for 13 campaigns.

The naive estimates show synergy that the estimates which account for differences in browsing types do not. The magnitude of the false synergy estimated here is not great, perhaps due to the

untargeted nature of these ads. However, I would expect to find a much larger spurious estimate for highly targeted ads where the synergy actually derives mechanically from the ad-targeting algorithms selecting users. That said, an absence of synergy is not necessarily bad—economically, the benefits of higher frequency are achieved as long as the marginal benefit of the additional clicks exceeds the marginal cost of delivering additional impressions.

#### 6.4. Heterogeneity and Wear-Out

Why is there such a wide range of varying ad performance? Perhaps some ads are better at being noticed. Alternatively, some ads may be more prone to wear out simply because they fail to appeal to a general audience. There are trade-offs between targeted messages that resonate with a particular segment of the population and fail to connect with the remaining viewers and a broader message that will not expire. It could be that the information conveyed by the ad is dated or time-sensitive such that visitors see the ads, find out the information, and move on after one impression. What creates such differences in frequency behavior across ads, and should advertisers be more concerned about the ads that wear out fast or those that fail to wear out?

Does the freshness of the creative matter? A number of the advertisers who experience decreasing returns to scale also experience some of the highest CTRs. From a basic review of the ad creatives, the content of these ads tended to be newer or more time-specific. In particular, a number of the ads experiencing the greatest wear-out were associated with new product releases (3 campaigns), new television episode (3 campaigns) or video releases (4 campaigns), or other time-sensitive or novel content (9 campaigns). The advertisers who experience constant returns to scale for their advertising exposure tended to display similar creatives (6 campaigns) or represent well-known brands for which there would arguably be relatively little new information coming from the ad (5 campaigns), other than a new opportunity to be reminded about the brand or product.

The wide variation in wear-out across advertisers may raise concerns about advertising effectiveness. However, advertisers should only worry about wear-out if the marginal effectiveness of the ad becomes sufficiently low for it to be unprofitable to advertise—not merely if an ad becomes less effective beyond the first exposure.<sup>7</sup> There is a trade-off between showing more ads to the best candidates and reaching marginal consumers who are less likely to respond. If the ads wear out slowly, then advertisers should not worry about reach, but should only target the most responsive subpopulation. If the ads wear out quickly, then advertisers should trade off the cost-ineffective frequency to reach more people, up to where the benefits to reach equal the benefits to frequency for the inframarginal reach and frequency.

<sup>7</sup> I forego discussing investment in new creatives for the sake of brevity. While ongoing investment in fresh new messages is encouraged by researchers such as Weilbacher (1970), I limit the advertiser's choice set "to whom should I target the ads" and "how many should be shown."

## 7. Analysis of Conversions

Campaign 10 was run by an advertiser with whom Yahoo! has a data-sharing partnership for anonymized conversion tracking. As a result, I can match users that were shown the advertiser’s ads on the Yahoo! Front Page and were shown a confirmation page upon completing the sign-up process for a new account on the advertiser’s website. This conversion data lets me measure the impact of frequency on this important outcome.<sup>8</sup> I use data from four campaigns run on four days.<sup>9</sup> Roughly 700 new account sign-ups occurred among the 40 million ad viewers each of the four days. I estimate the conversion model in equation 3 from section 4.1,

$$E[Y_{ic}|f, \theta] = \bar{b}_c \cdot b(\theta_{ic}) \cdot h(f_{ic}). \quad (7)$$

where  $h(\cdot)$  is no longer required to go through the origin, but through  $h(0) = 1$ . The relative frequency effects are displayed in Figure 9. In this figure, the relative frequency effects for sign-ups are in terms of the baseline sign-up rate when no ads are shown. The plot shows a large and positive impact of frequency from showing even as many as 20 impressions. The frequency impact on clicks for campaign 10 in Figure 11 are similar to those for sign-ups—both curves show relatively little wear-out after users view a large number of impressions in a single day. I compare the average effect of viewing twenty impressions on conversions with the effect of viewing only one. The estimated effect of one impression is 31% of the baseline and of twenty impressions is 627%, a ratio of 19.95.<sup>10</sup> Using the estimates of the frequency effects for clicks, the average effect of viewing the ads was 1.02 and 15.05 for one and twenty impressions, respectively, for a ratio of 14.79. These two ratios are statistically indistinguishable, indicating similar frequency effects for sign-ups as for clicks.

For the day following each campaign, the frequency impact on sign-ups was positive for small numbers of impressions, although the effects were statistically insignificant. However, the long-run impact of advertising on sign-ups is difficult to identify because the effect likely decays over time, reducing the signal-to-noise in estimation which is barely surmountable for the same day the ads are shown. Future research will pursue changes in both the estimation strategy and ad delivery that will improve the precision of these estimates.

The browsing heterogeneity curve in Figure 10 clearly deviates from the constant line of  $b(\theta) = 1$ , with heavy users being between two and three times more responsive to the advertiser’s offering

<sup>8</sup> These are not the only outcomes that the advertiser is interested in for campaign performance evaluation. Further, the conversion pixel being analyzed here only accounts for a fraction of the total number of sign-ups registered by users due to browser cookie deletion. As such, this analysis is primarily qualitative: does advertising frequency influence relevant outcomes such as account sign-ups?

<sup>9</sup> Each of the four campaigns was separated by at least a week from the other three campaigns.

<sup>10</sup> Note that this ratio is very imprecisely measured due to the volatility of the estimate of  $h(\cdot)$  for  $f = 20$  as seen in Figure 9.

than users who only visit the Yahoo! Front Page once. Ignoring this browsing type heterogeneity would produce naive frequency estimates that significantly overstate the impact of frequency on sign-ups for the four campaigns. The estimate of the causal effects of frequency on sign-up rates for the advertiser's four campaigns provides a much more reliable estimate of the impact of showing more ads to the same user.

## 8. Conclusion

I measured the impact of display advertising frequency for 30 campaigns shown on the Yahoo! Front Page between January and March 2010. The effect of display ad frequency on clicking behavior exhibits large heterogeneity across campaigns. Some advertisers find little decline in clicks to the marginal ad, even for as many as 30 to 50 impressions, while other advertisers experience rapidly decreasing clicks for showing even as few as two or three impressions. Of the 30 campaigns, ten (33%) experience constant returns to scale, six (20%) experience mild wear-out, ten (33%) experience moderate wear-out, and four (13%) experience extreme wear-out. For one advertiser, I can estimate the impact of frequency on new account sign-ups on the advertiser's website and find similar mild wear-out frequency effects for both sign-ups and clicks for that advertiser.

Carefully examining the frequency estimates, I find no convincing evidence for increasing returns to scale, commonly known as "ad synergy." However, naive estimates that ignore correlation of unobserved heterogeneity with the number of ads find false synergy for half of the 30 campaigns. Conversely, the naive estimator significantly overstates the wear-out of most of the 30 campaigns and understates the wear-out for a few. Whether or not an ad faces a declining marginal response rate, advertisers should show precisely the number of ads that equate the marginal benefit reaped from the last ad shown with its marginal cost. By better understanding the effects of frequency, better decisions can be made to deliver the right number of ads to the most responsive audience. Rather than merely seeking to "increasing reach," these methods can help advertisers reach the right customers the right number of times.

## Acknowledgments

I give special thanks to both the MIT Economics Department and Yahoo! Inc. for intellectual, financial, and data support and academic independence granted to complete this research. I thank Daniel Suh from the Yahoo! Front Page team for information on ad-splits. I appreciate the helpful comments from Justin Rao, David Reiley, Michael Schwarz, Dan Nguyen, Clara Lewis, Glenn Ellison, Jerry Hausman, Stephen Ryan, Nancy Rose, many participants at conferences and seminars, and anonymous referees. Thank you!

## References

Campbell, Margaret C, Kevin Lane Keller. 2003. Brand familiarity and advertising repetition effects. *Journal of Consumer Research* **30**(2) 292–304.

- Carlson, M., M. Hislop. 2001. 5 exposure plateau for building message online URL [www.dynamiclogic.com/beyond\\_1\\_8.html](http://www.dynamiclogic.com/beyond_1_8.html).
- Chang, Yuhmiin, Esther Thorson. 2004. Television and web advertising synergies. *Journal of Advertising* **33**(2) 75–84.
- Chatterjee, Patrali, Donna L Hoffman, Thomas P Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science* **22**(4) 520–541.
- Craig, C Samuel, Brian Sternthal, Clark Leavitt. 1976. Advertising wearout: An experimental analysis. *Journal of Marketing Research* 365–372.
- Dreze, Xavier, François-Xavier Husherr. 2003. Internet advertising: Is anybody watching? *Journal of interactive marketing* **17**(4) 8–23.
- Havlena, William, Robert Cardarelli, Michelle De Montigny. 2007. Quantifying the isolated and synergistic effects of exposure frequency for tv, print, and internet advertising. *JOURNAL OF ADVERTISING RESEARCH-NEW YORK-* **47**(3) 215.
- Johnson, Garrett, Randall A. Lewis, David H. Reiley. 2014. Location, location, location: Repetition and proximity increase advertising effectiveness. *Unpublished manuscript* .
- Kameya, Alison, Katherine Zmija. 2002. What makes online advertising effective?
- Lewis, Randall A., Justin M. Rao, David H. Reiley. 2011. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. *Proceedings of the 20th international conference on World wide web*. ACM, 157–166.
- Lewis, Randall A, David H Reiley. 2014. Online advertising and offline sales: Measuring the effects of retail advertising via a controlled experiment on yahoo! *Quantitative Marketing and Economics* 1–32.
- Mann, Don H. 1975. Optimal theoretic advertising stock models: A generalization incorporating the effects of delayed response from promotional expenditure. *Management Science* **21**(7) 823–832.
- Morgan Stanley Dean Witter. 2001. Does internet advertising really work? URL [www.morganstanley.com/techresearch/](http://www.morganstanley.com/techresearch/).
- Palda, Kristian S. 1965. The measurement of cumulative advertising effects. *The Journal of Business* **38**(2) 162–179.
- Pechmann, Cornelia, David W Stewart. 1988. Advertising repetition: A critical review of wearin and wearout. *Current issues and research in advertising* **11**(1-2) 285–329.
- Sahni, Navdeep. 2013. Advertising spillovers: Field-experiment evidence and implications for returns from advertising. *Working Paper* .
- Simon, Julian L. 1979. What do zielske’s real data really show about pulsing? *Journal of Marketing Research* 415–420.

Tellis, Gerard J. 1988. Advertising exposure, loyalty, and brand purchase: a two-stage model of choice. *Journal of marketing research* 134–144.

Weilbacher, William M. 1970. What happens to advertisements when they grow up. *Public Opinion Quarterly* 34(2) 216–223.

Zielske, Hubert A. 1959. The remembering and forgetting of advertising. *The Journal of Marketing* 239–243.

## Figures and Tables

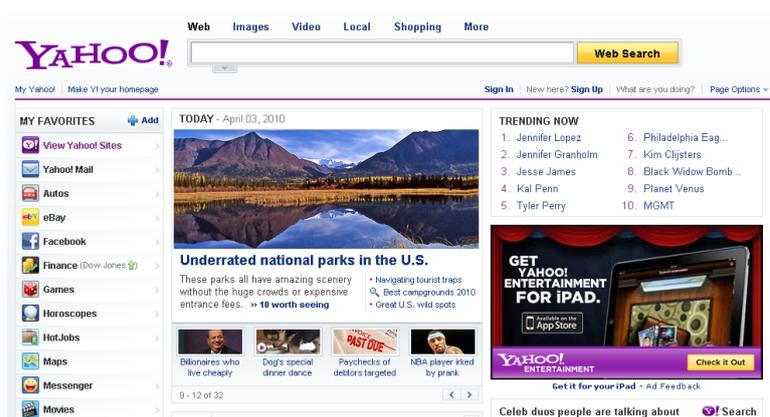


Figure 1 Example of Yahoo Front Page: April 3, 2010

**Table 1** Summary statistics for 30 Yahoo! Front Page “split” campaigns over 15 days.

Campaign	Impressions	Visitors	Imp./Visitor	Clicks	Clickers	CTR	% Clickers		
Day 1 {	20	98,177,379	41,221,308	4.76	2.38	51,653	46,010	0.05%	0.11%
	17	98,155,857							
Day 2 {	19	98,359,566	41,467,202	4.75	2.37	52,667	47,316	0.05%	0.11%
	10	98,428,392							
Day 3 {	25	98,718,046	41,993,478	4.70	2.35	50,004	44,633	0.05%	0.11%
	5	98,630,270							
Day 4 {	8	96,484,810	41,770,305	4.62	2.31	130,056	108,510	0.13%	0.26%
	9	96,576,838							
Day 5 {	18	94,700,877	40,611,100	4.67	2.33	54,142	48,665	0.06%	0.12%
	27	94,809,456							
Day 6 {	14	78,106,177	32,700,667	4.78	2.39	61,260	53,084	0.08%	0.16%
	30	78,157,493							
Day 7 {	3	91,620,276	41,811,276	4.38	2.19	284,565	253,503	0.31%	0.61%
	6	91,655,105							
Day 8 {	26	94,915,890	42,503,753	4.47	2.23	47,617	42,937	0.05%	0.10%
	21	94,924,172							
Day 9 {	7	70,325,235	31,925,737	4.41	2.20	104,522	87,509	0.15%	0.27%
	29	70,337,837							
Day 10 {	22	97,637,885	43,040,340	4.54	2.27	50,334	44,405	0.05%	0.10%
	11	97,596,427							
Day 11 {	4	95,612,568	42,513,766	4.49	2.25	179,173	158,226	0.19%	0.37%
	24	95,248,995							
Day 12 {	2	98,032,549	43,175,570	4.54	2.27	486,067	451,975	0.50%	1.05%
	23	98,053,868							
Day 13 {	1	93,770,718	42,221,664	4.44	2.22	659,665	575,327	0.70%	1.36%
	12	93,816,345							
Day 14 {	15	88,367,477	40,227,548	4.39	2.20	69,041	62,514	0.08%	0.16%
	13	88,299,108							
Day 15 {	28	94,742,664	42,092,579	4.50	2.25	46,686	41,691	0.05%	0.10%
	16	94,714,655							
<b>Totals &amp; Avgs.:</b>		2,778,976,935	609,276,293	4.56	2.28	3,507,228	3,110,801	0.13%	0.51%

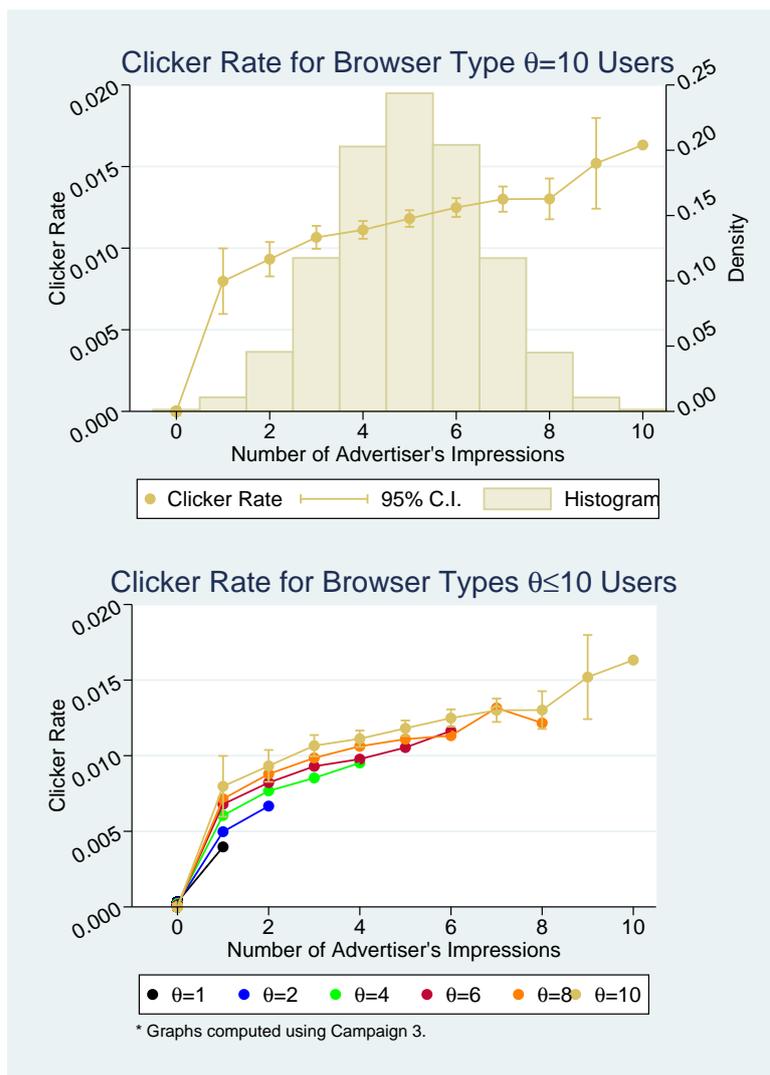


Figure 2 Click-through rate for users with type  $\theta \leq 10$  in campaign 3.

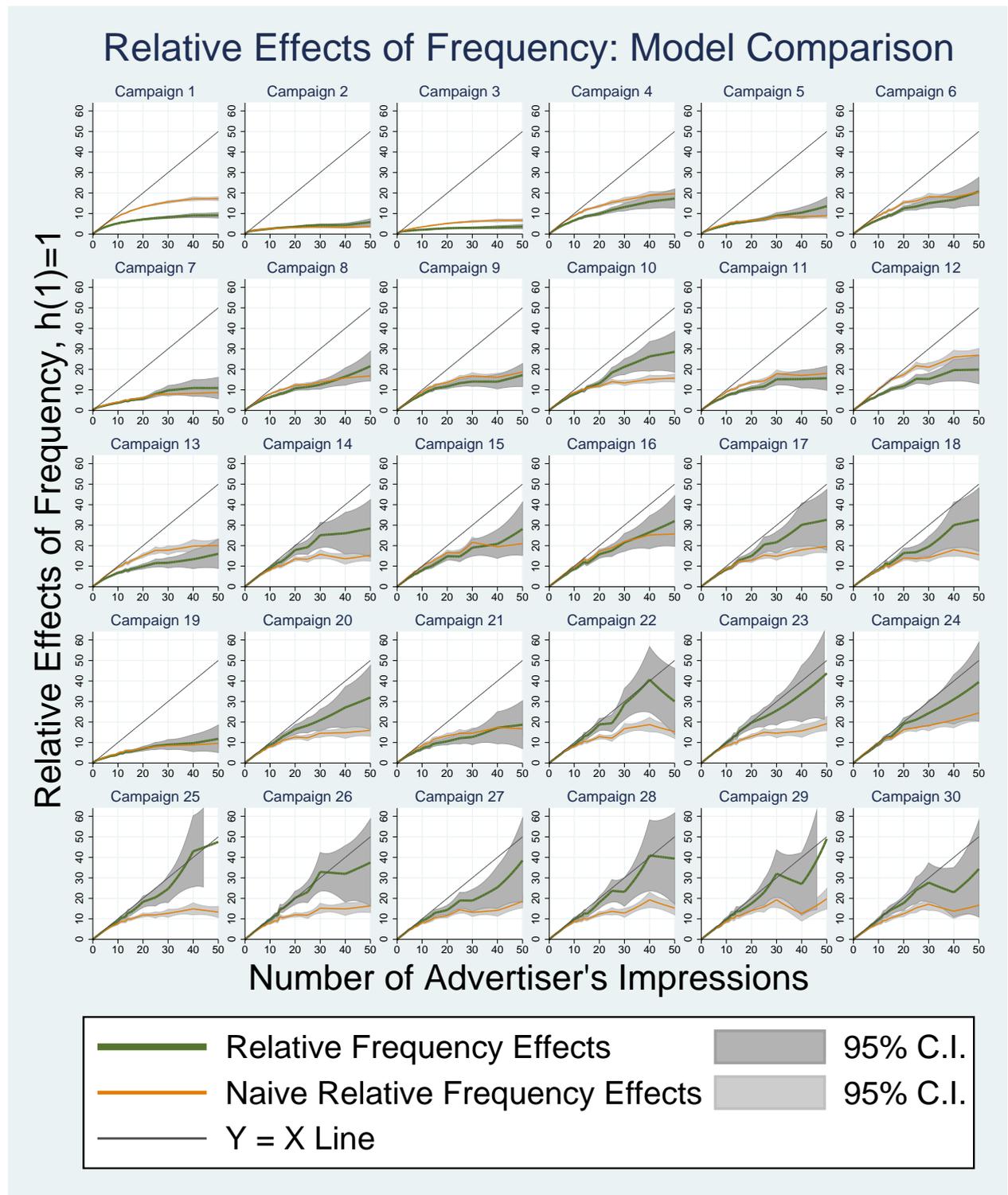


Figure 3 Relative frequency estimates with naive estimates for all 30 campaigns.

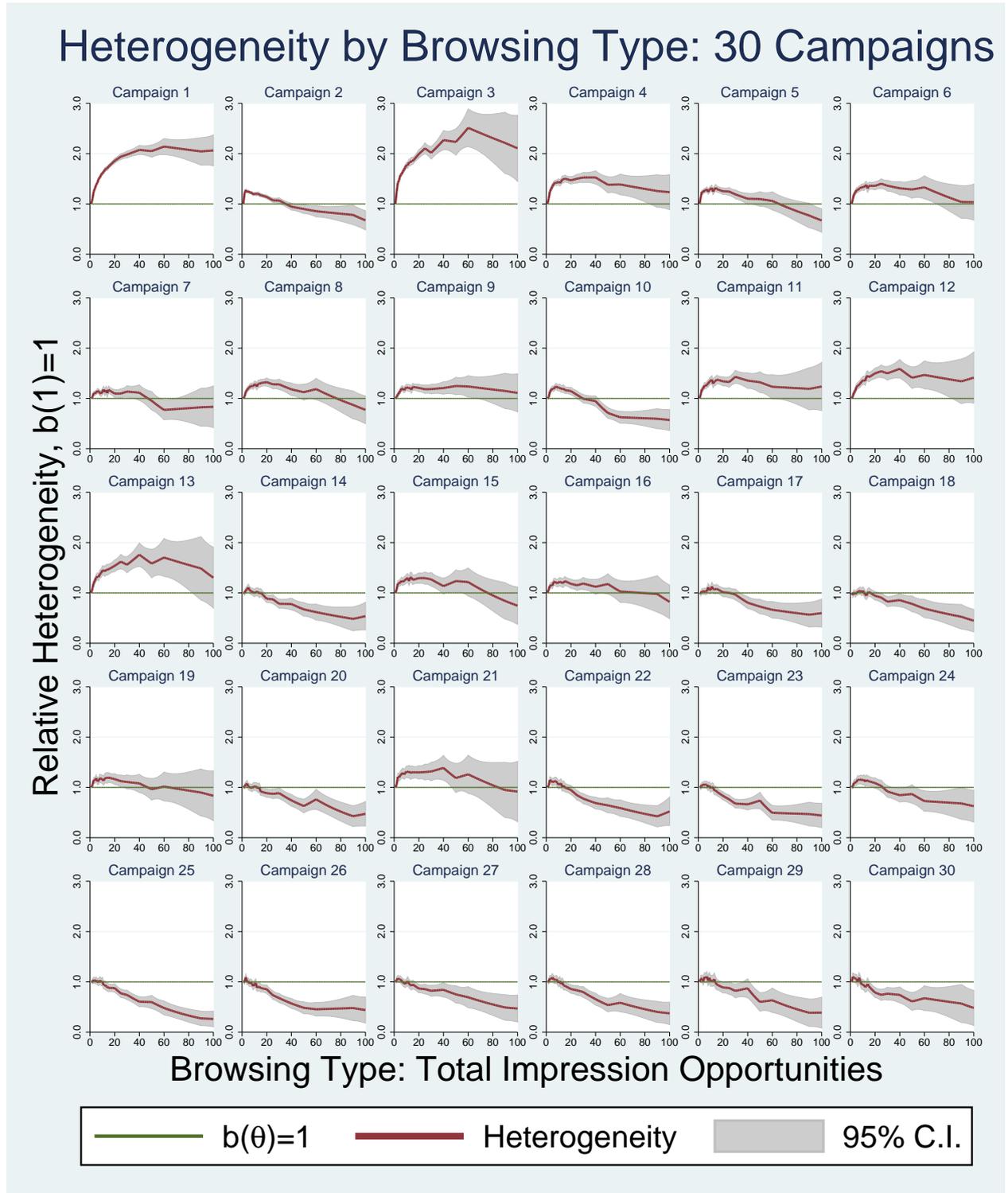


Figure 4 Heterogeneity by browsing type for all 30 campaigns.

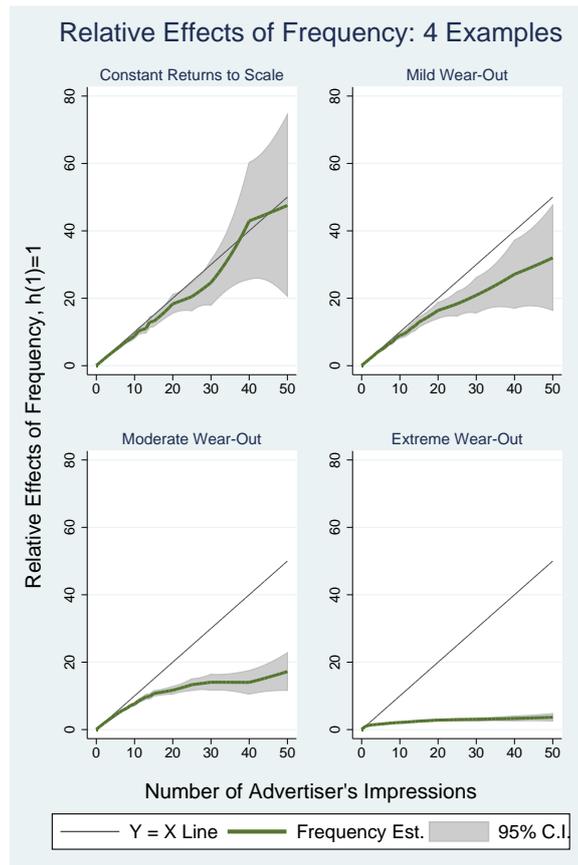


Figure 5 Four examples highlight the heterogeneous relative frequency response across campaigns.

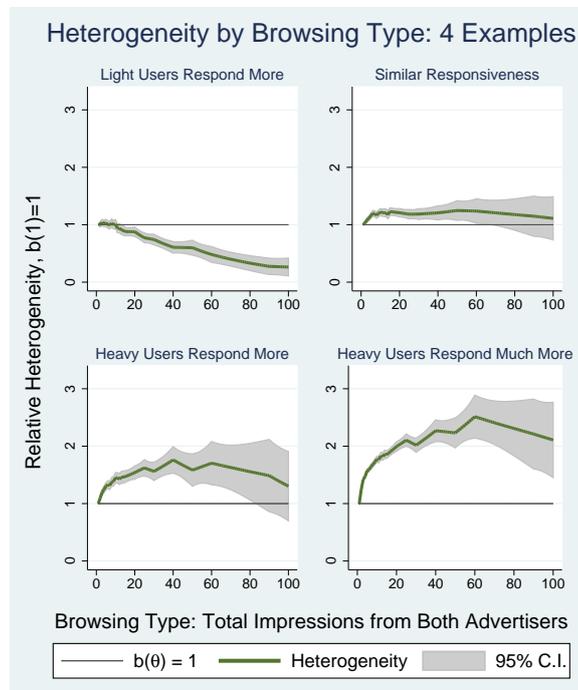


Figure 6 Four examples highlight the range of variation in responsiveness due to browsing type.

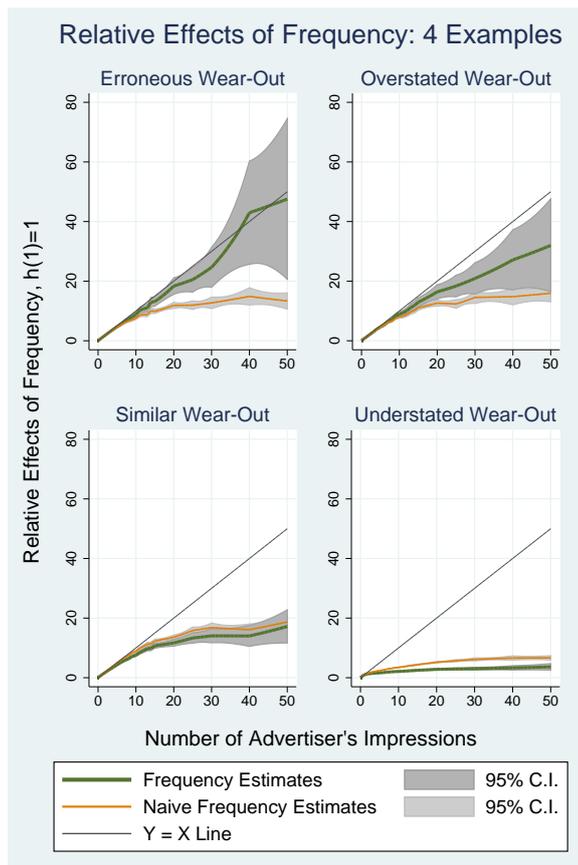


Figure 7 Naive frequency estimates incorrectly measure wear-out.

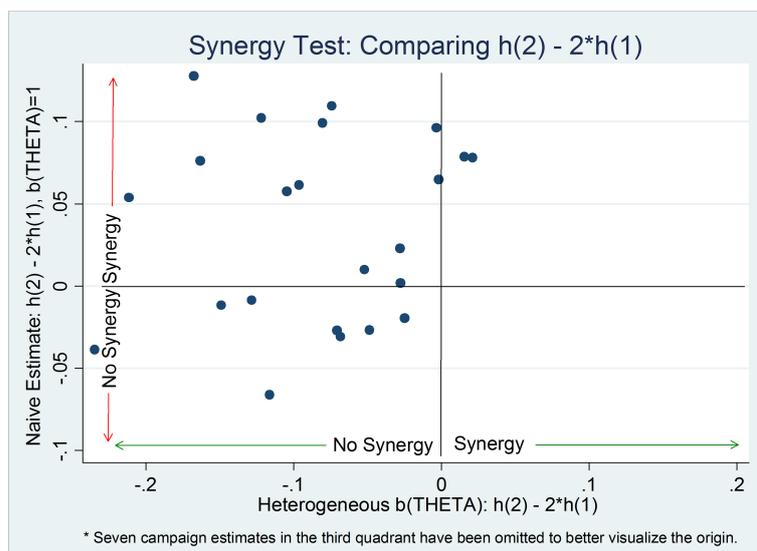


Figure 8 Naive frequency estimates find false ad synergy for many campaigns.

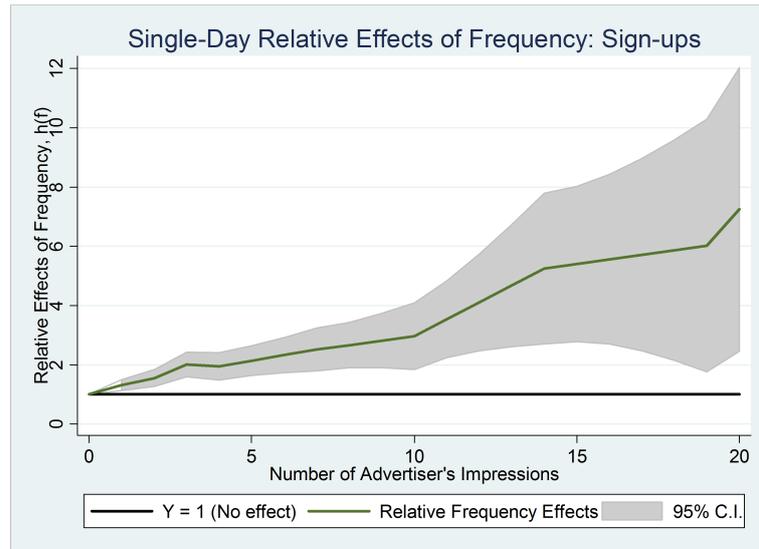


Figure 9 Relative effects of frequency on new account sign-ups for 4 campaigns.

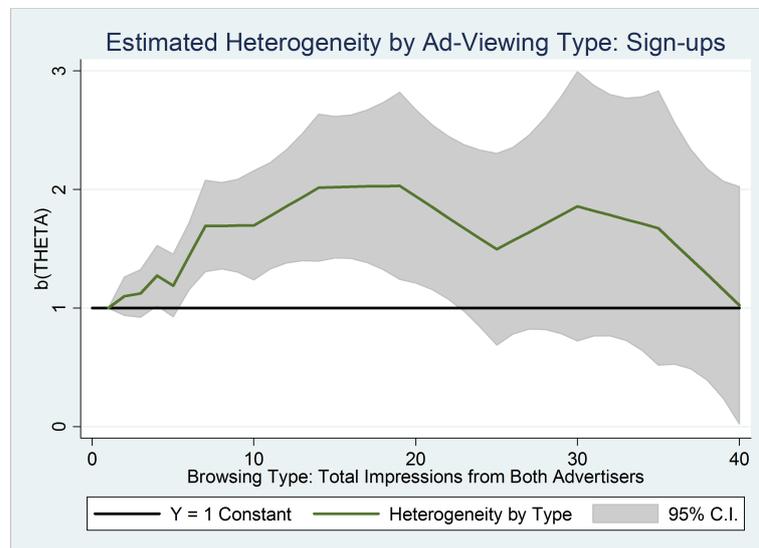


Figure 10 Relative heterogeneity for effects of frequency on new account sign-ups for 4 campaigns.

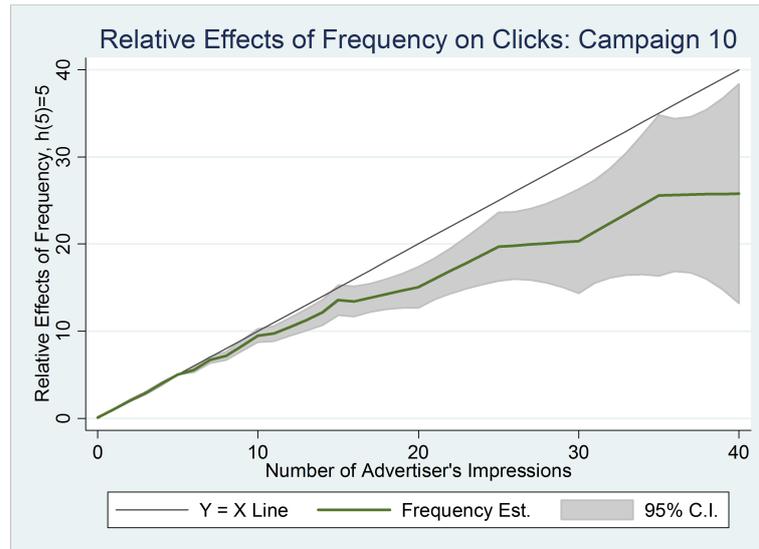


Figure 11 Relative effects of frequency on clicks for campaign 27.