A Large-Scale Field Experiment to Evaluate the Effectiveness of Paid Search Advertising

Lorenzo Coviello[*] Uri Gneezy[†] Lorenz Goette[‡]

September 19, 2017

Abstract

Companies spend billions of dollars online for paid links to branded search terms. Measuring the effectiveness of this marketing spending is hard. Blake, Nosko and Tadelis (2015) ran an experiment with eBay, showing that when the company suspended paid search, most of the traffic still ended up on its website. Can findings from one of the largest companies in the world be generalized? We conducted a similar experiment with Edmunds.com, arguably a more representative company, and found starkly different results. More than half of the paid traffic is lost when we shut off paid-links search. These results suggest money spent on search-engine marketing may be more effective than previously documented

[*]University of California San Diego, Department of Electrical and Computer Engineering, 9500 Gilman Drive, La Jolla, CA 92093 lorenzocoviello@gmail.com

[†]University of California San Diego, Rady School of Management, 9500 Gilman Drive #0553, La Jolla, CA 92093, ugneezy@ucsd.edu

[‡]University of Bonn, Institute for Applied Microeconomics, Adenauerallee 24-42, 53113 Bonn; lorenz.goette@uni-bonn.de

I Introduction

Advertising on search engines is a major expense for many companies. Companies spent an estimated 170 billion dollars on it in 2015. [1] Some observational studies simply count the traffic or revenue generated from paid search traffic (Ramos and Cota, 2008). However, this approach ignores the fact that some individuals would have found their way to the sponsored link anyway, thus potentially overstating the economic value of search-engine marketing. In a pioneering study, Blake et al. (2015) conducted a randomized trial to manipulate the availability of paid search advertising for so-called branded search involving eBay. These are searches that always include the 'brand' (eBay in this case) plus some other terms, such as 'eBay motorcycle.' According to their results, if eBay were to shut down its branded online advertisement, the volume of traffic to the eBay website would remain virtually unchanged, because traffic would still flow through the organic search results, costing no money to the advertiser.

Blake et al. (2015)'s conclusion is that their "evidence strongly supports the intuitive notion that for brand keywords, natural search is close to a perfect substitute for paid search, making brand keyword SEM [search engine marketing] ineffective for short-term sales. After all, the users who type the brand keyword in the search query intend to reach the company?s website, and most likely will execute on their intent regardless of the appearance of a paid search ad." These results are rigorous and valid in their own right, but are derived from the extreme case of eBay, which is one of the largest companies in the world.

In this paper, we test whether the results regarding branded ads from eBay, can be generalized to a firm of a more representative size in the industry. For instance, the top 1 percent of e-commerce companies are ranked within Alexa ranks 1 to 10,000, with a revenue share of approximately 34 percent of the market. But the lion?s share, 66 percent, of the revenues come from companies with distinctly lower Alexa ranks and, thus, visibility (Moore, 2014). In particular, can smaller companies increase the volume of traffic to their websites through paid brand search advertising?

We implemented a randomized controlled trial to estimate the effectiveness of branded search ads for Edmunds.com, a well-known online resource for automotive information (http://www.edmunds.com). Although Edmunds.com is not an extreme case like eBay, its Alexa rank (559 within the United States) places it well within the top one percent of most visited sites. With 700 employees and over $200 million in annual revenue, Edmunds.com is a large American company with strong online presence. However, it faces a more compet-

[1] Emarketer (2015) (https://www.emarketer.com/Article/Google-Will-Take-55-of-Search-Ad-Dollars-Globally-2015/1012294).

itive marketplace than eBay, with competitors of similar size, popularity, and services. We ask whether such companies can increase the volume of traffic to their websites through paid search advertising. Edmunds.com uses branded search-engine advertising in all of the 210 geographic markets in the United States, the so-called designated market areas (DMAs). For our study, we randomized half the markets, balanced by market size and penetration of Edmunds.com, into a control or a treatment condition. We monitored web traffic from August to November 2015 for each of these markets, being able to distinguish organic from paid traffic. After a baseline measurement period, branded search-engine advertising is shut off for the 105 markets in the treatment condition. This approach allows us to obtain a precise difference-in-differences estimate of the effect of branded search-engine advertising on overall web traffic and its components.

Our results are in striking contrast to those obtained by Blake et al. (2015) for eBay: only about half of the traffic normally accessing Edmunds.com through branded search ads still flowed to the website through organic search links. The remaining half likely landed on the pages of Edmunds' competitors who happen to bid on the keyword 'Edmunds.' The effect is particularly large in local markets with a high share of traffic from branded search ads in the baseline period. In these markets, Edmunds.com lost about 72% of this traffic as result of shutting down its paid brand search. That is, only 28% of branded search traffic still accesses the website via organic search links. Therefore, paid search advertising is far from ineffective even for a company as popular as Edmunds.

Note that in our intervention, we shut off branded search ads, arguably the most substitutable category in paid search advertising: typing the word 'Edmunds' in the search bar manifests a clear intention to visit a page on Edmunds.com. Whereas Blake et al. (2015) find 99.5% substitution for branded search-ads traffic for eBay, we find that substitution is less than 50% for Edmunds. Overall, our results suggest the findings by Blake et al. (2015) cannot be easily generalized even to a company within the top one percent of the most visited websites, such as Edmunds.com

Our paper is related to Chan et al. (2011), who identify naturally occurring temporal suspension of ads by companies and measure how much overall traffic decreases during these intervals compared to the previous periods. They report losses in overall web traffic of nearly 90% on average. However, their results are based on firms' decisions to suspend paid web traffic and are thus prone to endogeneity bias of various forms. For example, if firms anticipate low demand over a certain period, they may shut off search-engine marketing because they hold lower inventories during that period. This leads to overestimation of the effect of paid search advertising, as temporal variations in demand act as an omitted variable. This highlights again the importance of experimental variation in the availability of paid search-engine advertising in identifying its causal effect (also see, e.g., Lewis and Reiley, 2014; Lewis et al., 2011).

Simonov et al. (2015) manipulate the number of available paid ads on bing.com for branded searches. Bing allows a maximum number of four paid ads per search, and the authors report evidence from an experiment that exogenously manipulates this number to 3, 2, 1, or 0. On bing.com the original search term, for example, 'Edmunds Toyota,' would always list the ad by Edmunds.com at the top (as the first ad). Thus, the condition that most closely resembles our experiment is comparing one to zero paid ads. Going from one to zero ads means turning off Edmunds' branded ads, but also not allowing competitors to place their ads. Thus, this manipulation is different ours in which competitors are active.

The authors find much higher substitution rates than in our experiment: on average, only about 10 percent of the paid traffic is lost. This number is difficult to compare with our estimate, as we manipulate paid ads for the 'owner' of the keyword, while leaving the ads of potential competitors in place. By contrast, Simonov et al. (2015) identify the treatment effect of having no ads at all compared to having one's own ad only. The paper also complements our findings, because it systematically varies the number of competitors in paid add searches, while holding the target firm's ad constant. The study finds that a larger number of competitors reduces overall traffic towards the target firm, but the effect is relatively modest. By contrast, our findings show that, holding competitors constant, turning off one's own ad can lead to significant loss in web traffic.

The remainder of the paper is structured as follows: Section II describes the institutional background and experimental design we implemented. Section III discusses the results. Section IV concludes the paper.

II The Experimental Setup

In the following section, we present the results from a large-scale field experiment conducted at Edmunds.com in order to measure the effectiveness of branded search ads in terms of traffic to the website.

A. Institutional background

Edmunds is a well known online resource in the US automotive industry, located in Santa Monica, California. It provide buyers with a variety of services to retrieve information about dealers and offers for used and new inventory. As of 2016, it had 700 employees and revenue of over $100 million in 2016, and ranked 559th in the Alexa ranking of the most visited websites in the United States. Edmunds has a large number of competitors of comparable size and online presence, such as cars.com (over 1,000 employees, market cap over $1 billion, Alexa ranking 559 in the United States), autotrader.com (3,300 employees, revenue over $1 billion in 2016, Alexa ranking 360), and truecar.com (500 employees, revenue over $100 million in 2016, Alexa rank 1,543). Given its size and online presence, and the existence of competitors on the markets, Edmunds.com offers an interesting case study for the evaluation of the effectiveness of paid search advertising.

B. Paid brand search advertising

Branded search refers to queries to a search engine containing a brand keyword, "Edmunds" in our study (e.g., "Edmunds used Honda Civic 2014"). Non-branded search refers to queries that do not contain the word "Edmunds" (e.g., "Honda Civic 2014").

The results displayed on a search engine include paid search ads and organic (or unpaid) search links. Online advertisers pay the search engine for all impressions or clicks to their ads, but do not pay for organic search links. Importantly, search ads always appear at the top of the page, followed by organic search links (see Figure 1). The ranking of paid ads is determined by an algorithm that takes into account elements including the advertisers' bids. The ranking of organic search links is based on relevance and is determined by proprietary algorithms (e.g., Page Rank in the case of Google).

We reiterate that our study only considers branded search ads.

C. Experimental procedure.

The experiment closely follows the design in Blake et al. (2015). We conducted the experiment between August and November 2015, and limited it to Yahoo and Bing. We did not consider Google, because its relative share of branded search ads for Edmunds is low with respect to organic search traffic (about 1 percent of the volume) and would have required running the experiment for several months. By contrast, traffic volume from branded search ads for Yahoo and Bing accounts for 14 percent of the overall traffic volume. In total, we included 6,587 branded keywords containing the term "Edmunds" in the study.

Randomization occurred by geographical location, by assigning DMAs to one of two treatment groups (we henceforth refer to DMAs as markets). We assigned 105 markets to the control group, and 105 markets to the treatment group. We took several factors into account for the assignment of markets to treatment or control group: market condition (e.g., penetration rate of services offered by Edmunds.com), geographical location, market size (in terms of population and traffic to Edunds.com), and penetration of radio advertisement. Below, we show that markets in the treatment and control group do not differ statistically.

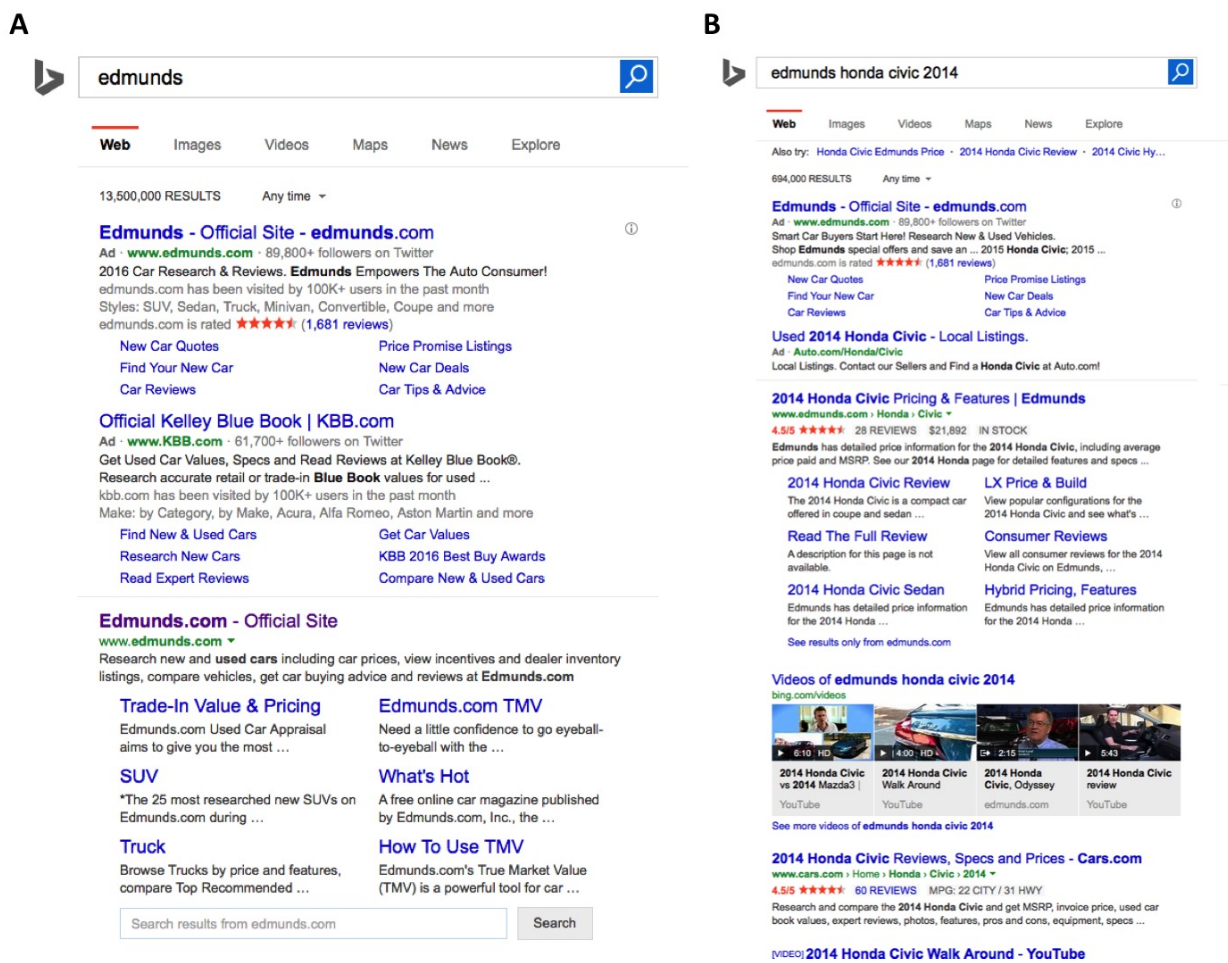

Figure 1: Examples of search results on Bing resulting from the search queries “edmunds” (panel A) and “edmunds honda civic 2014” (panel B), with paid branded ads at the top of the page followed by organic search links.

The experiment was characterized by two periods: a "baseline period" from August 9 to October 13, 2015, in which branded search ads were active as usual in all markets, both in the control and treatment groups; and an "intervention period" from October 14 to November 5, 2015, in which branded search ads from Yahoo and Bing were suspended in the treatment group DMAs.

For technical reasons, we could not suspend branded search ads on 100 percent of branded keywords terms, because different advertising campaigns run by Edmunds.com use different keywords and keyword matching criteria. [2] This resulted in residual traffic from branded search ads in the treatment-group markets during the intervention period. For each market and each day in the observation period, we consider the total number of sessions originated from organic search links and branded search ads on Bing and Yahoo, and for brevity refer to these quantities as the organic traffic and paid traffic, and to their sum as the total traffic. Note that the fact that paid traffic is only reduced, but not completely eliminated, does not impede us from estimating the degree of substitution, because we can compare what fraction of the estimated reduction in traffic is reflected in the change in total web traffic.

As a randomization check, we test whether the treatment and control groups differ statistically during the baseline period. Table 1 displays the results. In particular, we show that they do not differ in size and in the share of branded search traffic with respect to overall traffic. For each market, we consider the mean of daily traffic (paid plus organic) during the baseline period. In addition, we assign to each market the quantile value of its total amount of traffic during the baseline period (given the distribution of all markets). Finally, for each market, we compute the ratio between the average of daily paid traffic and the average of total traffic during the baseline period (average share of paid traffic during the baseline period). For each of the defined quantities, OLS regression shows that, in the baseline period, no significant differences exist between treatment and control markets (note that the constant term reflects the mean of the control group). Thus, randomization was successfully implemented.

[2] The main forms of keyword-matching criteria are Exact Match, Phrase Match, and Broad Match. Consider the keyword "Camry 2014 Edmunds." Exact match allows the bidder's ad to show only when the search query matches the keyword. Phrase match allows the bidder's ad to show when the search query is a close version of the keyword, with words before or after (e.g., "buy 2014 Camry Edmunds"). Broad match allows the bidder's ad to show when the search query is a variation of the keyword (e.g., "Edmunds used Toyota Camry").

Table 1: Randomization checks
OLS Regressions

Dependent variable: (measured during baseline phase)	Number of total daily sessions	Average quantile position	Fraction of Paid traffic
Treatment Market (=1)	14.885	-0.009	-0.004
	(32.931)	(0.040)	(0.004)
Constant	125.940***	0.507***	0.138***
	(21.564)	(0.027)	(0.003)
R^2	0.001	0.000	0.006
Obs	210	210	210

Notes: Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

III Results

A. Average traffic over time

Figure 2 shows aggregated, normalized traffic trends in the treatment and control markets during the period of observation. Given the difference in traffic volume between markets, daily traffic in each market (paid, organic, and total) is divided by its daily average total traffic during the baseline period. We refer to the resulting quantities as the normalized daily paid, organic, and total traffic in the market. The normalization factor for each market is computed over the baseline period only (so that it has the same interpretation for both control and treatment markets), and the same normalization factor is used for paid, organic and total traffic, so that the shares of normalized paid and organic traffic add up to 1 in the baseline period.

As the figure shows, the time trends in treatment and control markets are the same during the baseline period.

The left panel shows the normalized volume of paid traffic for both treatment and control markets in the baseline period, which is about 12 percent to 15 percent of overall traffic. It also shows that the intervention almost completely shuts off paid traffic in treatment markets, reducing it to at most 2 percent to 3 percent of baseline total traffic. As mentioned above, paid traffic was not completely shut off, due to the coexistence of other experiments performed by Edmunds.

The middle panel shows that during the intervention period the volume of organic traf-

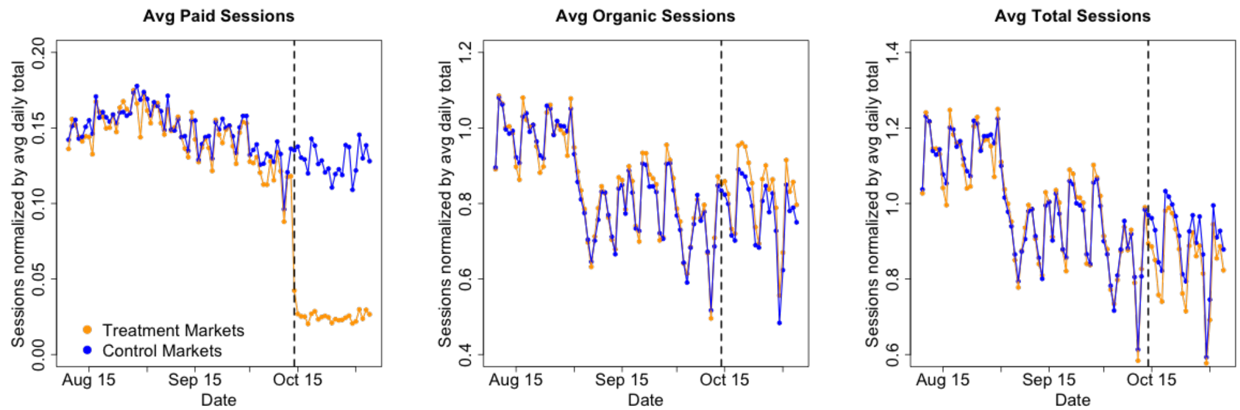

Figure 2: Time trends of normalized traffic volume over time, originated by branded search ads (left panel), organic search links (middle panel), and total (right panel), averaged over control markets (blue) and treatment markets (orange).

fic in treatment markets increases with respect to control markets, in a potential sign of substitution of organic and paid traffic. However, the increase in organic traffic does not fully compensate the reduction in paid traffic, as shown by the right panel: the normalized overall traffic in treatment markets during the intervention period appears to decrease compared to control markets.

B. Traffic change during the intervention period

As a first descriptive step towards a full difference-in-differences estimate of the treatment effect, we proceed to a descriptive analysis of the change in web traffic for treatment and control markets between the baseline period and the intervention period.

Denote the total traffic in market i on day t as y_{it}. Denoting paid and organic traffic as y_{it}^{paid} and y_{it}^{org}, we have that $y_{it} = y_{it}^{paid} + y_{it}^{org}$. The average total traffic in market i during the baseline and the intervention period can be written respectively as

$$y_{i,0} = \sum_{t \in T_0} y_{it}, \quad y_{i,1} = \sum_{t \in T_1} y_{it},$$

where T_0 and T_1 denote the baseline and intervention periods. $y_{i,0}^{paid}$ and $y_{i,1}^{paid}$ are similarly defined for paid sessions, and $y_{i,0}^{org}$ and $y_{i,1}^{org}$ for organic sessions.

The average change in traffic volume in market i between the intervention period and the baseline periods is therefore $y_{i,1} - y_{i,0}$, $y_{i,1}^{paid} - y_{i,0}^{paid}$, $y_{i,1}^{org} - y_{i,0}^{org}$ for total, paid, and organic session respectively. Considering these differences rather than a time series for each market

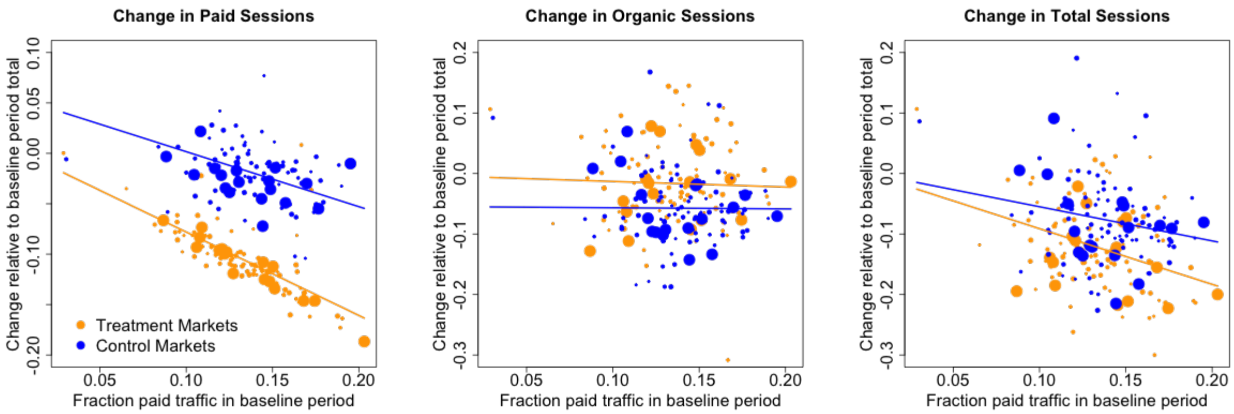

Figure 3: Normalized change of traffic from the baseline to the intervention period in each market. Traffic originated by branded search ads (left panel), organic search links (middle panel), and total (right panel). Control markets are represented in blue and treatment markets in orange. Points size is proportional to market size (in terms of overall traffic in the baseline period). Lines show least squares fit (non-weighted by market size).

eliminates the effects of time trends, and allows to focus on the impact of the intervention. To control for differences in traffic volume between markets we consider the normalized changes in traffic volume, dividing by the average total traffic during the baseline period $y_{i,0}$:

$$\Delta y_i = \frac{y_{i,1} - y_{i,0}}{y_{i,0}}, \quad \Delta y_i^{paid} = \frac{y_{i,1}^{paid} - y_{i,0}^{paid}}{y_{i,0}}, \quad \Delta y_i^{org} = \frac{y_{i,1}^{org} - y_{i,0}^{org}}{y_{i,0}}.$$

Note the all three quantities are normalized by $y_{i,0}$ (baseline total traffic in the market), and therefore $\Delta y_i = \Delta y_i^{paid} + \Delta y_i^{org}$ by construction.

Figure 3 plots the normalized traffic change in the intervention period in each market. In particular, the variables Δy_i^{paid}, Δy_i^{org}, and Δy_i are plotted against the fraction of paid traffic in market i during the baseline period:

$$f_i = \frac{y_{i,0}^{paid}}{y_{i,0}}.$$

The left, middle and right panels display the change paid, organic, and total traffic, respectively, normalized by the size of the market in the baseline phase. We also superimpose OLS estimates of the change in web traffic as a function of the fraction of paid traffic in the baseline phase for control (blue) and treatment markets (orange). From the left panel, the blue dots show regression to the mean effect for control markets: markets with a higher share of paid traffic during the baseline period (higher f_i) show a higher reduction in paid traffic in the intervention period with respect to markets with a lower share.

The orange dots, representing the change in paid traffic in the treatment markets, shows a markedly larger reduction in paid traffic. Most of the orange points are below the blue ones (for a given fraction of paid traffic in the baseline period), and the slope with respect to the fraction of paid traffic is visibly steeper than in the control markets.

As we mentioned before, paid traffic was not shut off completely, even in the treatment markets. if paid traffic were shut off completely in a market, the corresponding dot would all lie on a downward diagonal line with slope -1 and intercept 0). The graph suggests our experiment shut off about 75 percent of paid traffic in treatment markets. Overall, the experimental intervention is a strong manipulation in paid traffic that allows to study how organic traffic responds to it.

The middle panel displays the change in organic traffic. Treatment markets show an increase in organic traffic with respect to control markets: the orange dots (and the corresponding OLS fit) are above the blue ones. This effect is expected under the substitution hypothesis that, when branded search ads are shut off, (part of) the corresponding traffic becomes organic traffic.

However, the figure also suggests the increase in organic traffic does not offset the entire drop in paid traffic. This can be seen in the right panel, which displays the results for total traffic. Although treatment and control markets present a reduction in traffic volume from the baseline period to the intervention period, the reduction in traffic is substantially larger in treatment markets (the orange dots and line are below their blue counterparts). That is, the increase from organic traffic in control markets did not offset the loss of paid traffic. In addition, the negative slope of the orange line shows that markets with a higher share of paid traffic during the baseline period experienced a larger loss in overall traffic volume. We do not find that complete substitution of paid traffic through organic traffic occurs, in sharp contrast to the observations by Blake et al. (2015) for eBay.com.

C. Difference-in-differences estimates

To obtain a formal estimate of these effects, we now turn to a regression framework, with the following model specification:

$$\Delta y_i = \beta_0 + \beta_1 T_i + \varepsilon_i, \tag{1}$$

$$\Delta y_i = \beta_0 + \beta_1 T_i + \beta_2 f_i + \varepsilon_i, \tag{2}$$

where T_i is the treatment indicator equal to 1 for treatment markets and 0 for control markets, and f_i is the fraction of paid traffic in the baseline period in market i. We consider similar models for Δy_i^{paid} and Δy_i^{org}. Note that we consider a single observation for each market, avoiding complications from potential serial correlation in the time series. The parameter of interest is β_1, which represents the impact on the normalized change in traffic of shutting off branded search ads. /footnotestrictly speaking, this is the effect of *intending* to shut off branded search ads, given that paid traffic was not completely shut off during the intervention, which results in more conservative estimates of the treatment effect. Including the fraction of paid traffic in the baseline substantially increases the precision of the estimates (increase in R^2, decrease in standard errors).

To calculate valid standard errors, we need to address two problems: first, the experiment could change the variance in markets. Thus, we need to estimate heteroskedasticity-robust standard errors. Second, the difference in scale in markets still affects the variance of our residuals. Intuitively, large markets will provide a more precise estimate of the percentage change (because the numerator is larger, and, in relative terms, less volatile). Thus, the variance of the residual is inversely proportional to baseline traffic. We correct for this by using weighted least squares with regard to the market size in the baseline period.

Table 2 displays the results. The two leftmost columns shows the impact on paid traffic Δy_i^{paid}. Considering model specification (1) in the first column, paid traffic is reduced by 9.8 percentage points (expressed in units of overall traffic in the baseline period) in treatment markets. Given an average share of paid traffic of 14% in the baseline period, this shows that the intervention drastically reduced paid traffic (even if it did not shut it off completely). Model specification (2) in the second column gives much higher precision and a larger estimate of the treatment effect. Moreover, the coefficient for f_i confirms the regression to the mean effect mentioned above, according to which markets with a larger share of paid traffic experienced larger decrease in paid traffic in the intervention period. Because this estimate eliminates the variance generated by regression to the mean, the estimate of the treatment effect is notably more precise.

The two middle columns show the impact on paid traffic Δy_i^{org}. Considering model specification (1), the estimates in the third column show treatment markets experience an increase in organic traffic of 4 percentage points (expressed in units of overall traffic in the baseline period) relative to control markets, and the effect is highly significant. However, this effect is qualitatively far smaller than the one observed by Blake et al. (2015).

The two rightmost columns shows the impact on overall traffic Δy_i. Considering model specification (1) in the fifth column, treatment markets loose approximately 5.6 percentage points in traffic volume relative to control markets. Model specification (2) in the sixth columns give much higher accuracy, and a slightly higher estimate of 6.2 percentage, In short, the increase in organic traffic was not able to compensate for the loss in paid traffic,

Table 2: Difference-in-differences estimates of the treatment effects
WLS Regressions

Dependent variable: change in web-traffic category, normalized by average total web traffic in market during the baseline phase.

Dependent variable:	paid traffic		organic traffic		total traffic	
Treatment Market (=1)	-0.098***	-0.102***	0.042***	0.040***	-0.056***	-0.062***
	(0.008)	(0.003)	(0.012)	(0.011)	(0.017)	(0.012)
Fraction of paid sessions in BL		-0.756***		-0.435*		-1.191***
		(0.100)		(0.255)		(0.321)
Constant	-0.020***	0.092***	-0.077***	-0.012	-0.097***	0.080
	(0.002)	(0.015)	(0.009)	(0.040)	(0.009)	(0.050)
R^2	0.746	0.918	0.163	0.232	0.173	0.473
Obs	210	210	210	210	210	210

Notes: Heteroskedasticity-robust WLS standard errors are in parentheses. Estimates are weighted by the average total web traffic in a market during the baseline (the normalizing variable). *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

in contrast to the results by Blake et al. (2015) for eBay.

D. Difference-in-differences estimates in subgroups

Next, we quantify how the heterogeneity of the markets, in terms of their dependence on paid traffic during the baseline period f_i affects the impact of shutting off branded search ads. Differences across markets in the reliance on paid searches may be correlated with how "active" or "passive" consumers are in their searches. If use of paid search ads is indicative of passive search behavior by consumers, we would expect markets with a higher fraction of paid searches in the baseline phase to have a larger overall loss in web traffic, because these customers may be more easily captured by competing paid ads that show up conveniently at the top of the screen (see, e.g., Fowlie et al., 2015, for an analysis of how passive consumers respond less to the economic environment they are facing).

We thus interact the treatment indicator T_i with the fraction f_i (reliance on paid traffic in the baseline period), and estimate the following model:

$$\Delta y_i = \beta_0 + \beta_1 T_i + \beta_2 f_i + \beta_{12} T_i \times f_i + \varepsilon_i,$$

as well as analogous models for Δy_i^{paid} and Δy_i^{org}. Again, we report heteroskedasticity-robust weighted least squares estimates as in the previous table.

Table 3: Difference-in-differences estimates of subgroup analyses
WLS Regressions

Dependent variable: change in web-traffic category, normalized by average total web traffic in market during the baseline phase.

Dependent variable:	paid traffic	organic traffic	total traffic
Treatment Market (=1)	0.008	0.088	0.096
	(0.009)	(0.077)	(0.078)
TM x Fraction paid in BL	-0.748***	-0.327	-1.075**
	(0.061)	(0.491)	(0.494)
Fraction of paid sessions in BL	-0.226***	-0.203	-0.428
	(0.052)	(0.392)	(0.377)
Constant	0.013*	-0.047	-0.033
	(0.007)	(0.062)	(0.061)
F-test: no impact of treatment	$p < 0.001$	$p < 0.001$	$p < 0.001$
Predicted treatment effects for markets			
with 10% paid traffic in BL	-0.07***	0.054***	-0.016
	(0.003)	(0.027)	(0.028)
with 17% paid traffic in BL	-0.118***	0.032***	-0.086***
	(0.002)	(0.012)	(0.012)
R^2	0.953	0.240	0.524
Obs	210	210	210

Notes: Heteroskedasticity-robust WLS standard errors are in parentheses. Estimates are weighted by the average total web traffic in a market during the baseline (the normalizing variable). *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

The first column in Table 3 shows how the intervention affects paid traffic Δy_i^{paid} depending on markets' reliance on paid traffic during the baseline period. The point estimate of the main treatment, which represents the treatment effect for a market with zero paid traffic in the baseline period, is virtually zero, which is expected. The interaction term quantifies the reduction of paid search traffic during the intervention for a one-percent point increase in paid traffic during the baseline phase. The point estimate of -0.748 indicates that a one-percentage-point increase in baseline paid traffic leads to a loss of approximately three quarters of a percentage point in paid traffic during the intervention. As explained above, branded search ads were not entirely shut down; therefore the reduction is less than one for one. Nevertheless, markets with a higher baseline fraction of paid traffic lose much more traffic in response to our experiment.

By using both the main effect and the interaction effect, the bottom panel of the Table 3 (first column) quantifies the predicted treatment effect for markets at the 10th and 90th percentile of the share of paid traffic during the baseline period. The two cases can be considered as examples of markets with low and high reliance on paid traffic respectively. Markets at the 10th percentile had approximately 10 percent of paid traffic during the baseline period, and our model predicts a loss of about 7 percentage points (of the 10 percent share in the baseline period) during the experiment. The effect on markets with a higher reliance on paid traffic is substantially stronger: markets at the 90th percentile of baseline paid traffic had a share of about 17 percent of paid traffic in the baseline period, which corresponds, according to our estimates, to a loss of nearly 12 percentage points (out of the 17 total) during the experiment.

The second column of Table 3 considers the effects on organic traffic, Δy_i^{org}. Interestingly, the point estimate is negative, suggesting markets with a higher reliance on paid traffic experienced a weaker increase in organic traffic, even though they clearly lose more in terms of paid traffic. However, the coefficient estimate is not significant. Overall, organic traffic increased during the intervention for the treatment markets, as shown in Table **??**, and as confirmed by an F-test testing the joint hypothesis that both the main effect and the interaction effect are zero, rejected at a 0.001 confidence level (see bottom panel of the Table 3, second column). However, our estimates lack the precision to determine whether the shift is uniform, or whether it scales with the baseline share of paid traffic.

As above, the bottom panel of the Table 3 (second column) reports the predicted treatment effects on organic traffic for markets at the 10th and 90th percentile of baseline paid traffic. The point estimates show no evidence that markets losing more paid traffic experience a stronger increase in organic traffic.

The third column of Table 3 considers the effects on overall traffic, Δy_i. While, not surprisingly, the imprecision of organic traffic carries over to overall traffic, the interaction effect with the baseline share of paid traffic is significant. The point estimate of -1.08 suggests that a one-percentage increase in the baseline share of paid traffic leads to a complete incremental loss of that percentage point during the experiment.

The bottom panel of Table 3 (third column) again shows the predicted treatment effect for markets at the 10th and 90th percentiles of baseline paid traffic. For markets with low reliance on paid traffic (of approximately 10 percent of the overall traffic in the baseline period), the overall loss is very small, amounting to 1.6 percent of traffic, and is not significantly different from zero. By contrast, our model predicts a high loss in traffic for markets with a high reliance on paid traffic. A market at the 90th percentile in paid traffic share (about 17% of total traffic) loses a sizable amount of overall traffic: of the 11.2 percentage point reduction in paid traffic, 8.6 percentage points are lost when shutting off branded search ads. In other words, 72 percent of the paid traffic is lost for markets with a high baseline share of paid traffic.

IV Discussion and Conclusion

We conducted a randomized controlled trial to estimate the effectiveness of online branded search ads for Edmunds.com, a well-known online resource for automotive information. In our experiment, we shut down online brand advertisement in some regions, and compared the change in traffic with the change in traffic in control regions. We ask how this shutdown affects the overall volume of traffic to Edmunds.com , and what fraction of traffic is simply substituted through the unpaid organic search.

Our results show that shutting off branded search ads results in a net loss of about 50 to 60 percent of the traffic volume that normally accesses Edmunds.com via branded search ads. These results are in striking contrast to those obtained in the pioneering study of Blake et al. (2015). They found almost complete substitution of paid traffic with organic traffic when testing the effect of these ads on eBay. Our findings suggest the results in Blake et al. (2015) are limited to giant companies such as eBay, and cannot be generalized to most of the market – not even to large companies with a strong online presence such as Edmunds.com.

Interestingly, markets with a larger share of paid traffic before the intervention experienced lower substitution of paid traffic with organic traffic in response to our experiment. In markets with the highest reliance on paid traffic, less than 30 percent of paid traffic reverted to the organic channel, whereas in markets with the lowest reliance on paid traffic, about 80 percent reverted to the organic channel. This result opens a new research question regarding who uses paid traffic for different types of websites. For example, higher reliance on paid traffic might reflect the presence of less sophisticated consumers, who tend to click the links at the top of the search-results page. Therefore, firms or industries with a large share of such unsophisticated consumers might benefit more from paid search ads, because substitution is exceptionally low in these markets.

Estimating returns on investment for advertising is inherently difficult (Lewis and Rao, 2015). The study by Blake et al. (2015) stands out because of its precise measurements of returns. For Edmunds.com, most of the revenue is generated directly by the number of people who visit its website, for example, by selling ad-impressions to auto-related products. Based on the data we generated, Edmunds.com calculated that the revenue generated by the paid branded ads was significantly higher than the cost of the ads, and hence decided to continue with this operation.

Our message is clear: although the largest companies online, such as eBay, may lose money investing in paid brand search, many others may profit from it. A plausible reason for the differential return on such ads between companies relates to the nature of competition in the market. For example, Simonov et al. (2015) find the number of competing paid links plays a crucial role in diverting branded search traffic. Companies should invest in understanding the value of paid brand-ads for profits in their relevant domains.

References

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

Chan, David X., Yuan Yuan, Jim Koheler, and Deepak Kumar, "Incremental Clicks: The Impact of Search Advertising," *Journal of Advertising Research*, 2011, *51* (4), 643.

Fowlie, Meredith, Catherine Wolfram, C Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers, "Default Effects, Follow-on Behavior and Welfare in Residential Electricity Pricing Programs," Technical Report, Working Paper 2015.

Lewis, Randall A and David H Reiley, "Advertising effectively influences older users: How field experiments can improve measurement and targeting," *Review of Industrial Organization*, 2014, *44* (2), 147–159.

_ **and Justin M Rao**, "The unfavorable economics of measuring the returns to advertising," *The Quarterly Journal of Economics*, 2015, p. qjv023.

_ **,** _ **, and David H Reiley**, "Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising," in "Proceedings of the 20th international conference on World wide web" ACM 2011, pp. 157–166.

Moore, Robert J., "How Many Ecommerce Companies Are There?," 6 2014.

Ramos, Andreas and Stephanie Cota, *Search engine marketing*, McGraw-Hill, Inc., 2008.

Simonov, Andrey, Chris Nosko, and Justin M. Rao, "Competition and Crowd-out for Brand Keywords in Sponsored Search," Technical Report, Columbia University 2015.