Firms' Reactions to Public Information on Business Practices: The Case of Search Advertising*

Justin M. Rao
HomeAway, Inc.

Andrey Simonov
Columbia University

February 24, 2018

Abstract

We use five years of bidding data to examine the reaction of advertisers to widely disseminated press on the lack of effectiveness of brand search advertising (queries that contain the firm's name) found in a large experiment run by eBay (Blake, Nosko and Tadelis, 2015). We estimate that 11% of firms that did not face competing ads on their brand keywords, matching the case of eBay, discontinued the practice of brand search advertising. In contrast, firms did not react to the information pertaining to the high value and ease of running experiments—we observe no change in the experiment-like variation in advertising levels. Further, while 72% of firms had sharp changes in advertising suitable for estimating causal effects, we find no correlation between firm-level advertising effects and the propensity to advertise in the future. We discuss how a principal-agent problem within the firm would lead to these learning dynamics.

*Rao: justinmrao@outlook.com; Simonov: asimonov@gsb.columbia.edu. We thank Matthew Gentzkow and Matt Goldman for useful comments and suggestions. All opinions represent our own and not those of our current or past employers. All remaining errors are our own.

1 Introduction

Empirical findings in economics and management science often have normative implications for firm behavior. Papers that speak directly to the profitability of decisions faced by firms appear regularly in peer-reviewed journals. Particularly relevant findings are amplified by media coverage. Perhaps surprisingly, then, very little is known about how scientific progress in these areas impacts business practice. Our lack of knowledge can be explained by a few factors. First, it can be difficult to record decisions with the required granularity and quantitative rigor. Second, it is generally even more difficult to link these choices to their direct consequences. Third, even when such measurement is possible, the relevant data usually do not enter the academic discourse or do so via one-off arrangements with single firms, making it hard to draw broader conclusions about how firms learn.

We aim to overcome these challenges with a unique combination of detailed data on firm-level decisions with directly measurable consequences and the release of a particularly impactful academic paper that received widespread media attention. Our domain is sponsored search, a popular form of online advertising in which advertisers bid on slots at the top of search engine results. Since an advertiser can be present in both the sponsored links and the "organic" links returned by the search engine, there is the possibility of crowd-out between paid and free clicks. Blake, Nosko and Tadelis (2015, herein BNT) study this crowd-out with a field experiment involving tens of millions of dollars of search advertising. The experimental results show that ads on branded keywords (e.g. "eBay shoes") had almost no effect on traffic to eBay. Despite the minuscule causal impact, the nominal metrics of these ads—high click-through-rate (CTR) and low cost-per-click (CPC)—made them appear to be strong performers. Based on these results, eBay stopped advertising on branded terms, recouping an annual expenditure exceeding fifty million dollars.

BNT's results seemingly call into question the entire enterprise of advertising on own-brand queries. Since the practice is widespread, it is no surprise that the paper received attention in the popular press (*e.g. The BBC*), search engine marketing blogs (*e.g. Search Engine Land*) and the business press (*e.g.* the *The Economist*). We study the impact of this information disclosure using detailed bidding data covering a five-year period on the Bing search engine, focusing our analysis on the roughly one thousand firms for which there is adequate data coverage. We examine firms' reactions in two dimensions: 1) the propensity to advertise on own-brand queries; 2) the propensity to conduct experiments to measure ad effectiveness.

A natural way a firm could react to BNT's findings is to judge whether or not they are in a comparable situation as eBay and, if so, simply follow the example of eBay by stopping brand search advertising. A critical part of this "comparable situation" caveat is that eBay did not, during the time period of study, face competing advertisers that stood to supplant them in the top slot.[1] This is noted as a critical consideration, and Simonov, Nosko and Rao (2017) subsequently showed that the value of brand ads indeed lies in the ability to prevent a competitor from occupying the top spot.[2] Based on this observation, we use a difference-in-differences framework specifying firms that did not regularly face competitors as the "treated" group and firms that did regularly face competition as the "untreated" group. We find that treated firms decrease advertising levels on brand keywords by 11% relative to their untreated counterparts. The untreated companies show no break in the pre-existing time-trend, whereas treated companies show a marked downward shift in the propensity to advertise. The result is significant beyond the 0.01 level and robust to functional form assumptions, changes in the treatment assignment threshold and various other robustness checks.

The relatively modest impact on the propensity to advertise could be driven by the heterogeneity of brand search advertising effectiveness; perhaps only the firms that are similar to eBay tend to react. We do not find any detectable difference in the reaction of more prominent firms for which brand search advertising should be less effective and the results of BNT more relevant (Simonov et al., 2017). However, we find that firms for which own-brand advertising serves as a deterrent of competitors' entry tend to be less likely to react to BNT by stopping to advertise. Indeed, for a subset of companies, focal brand advertising deters competitors from entry; on average, competitors are 8.4 percentage points more likely to advertise when the focal brands do not, with a substantial heterogeneity across the focal brands.

Another explanation for the modest impact of BNT on the propensity to advertise is that,

[1] The ads above the organic links account for the vast majority of search engine revenue.

[2] Simonov et al. (2017) estimates that competitors on average "steal" around 18 percent points of clicks when they are in the top position (first paid link), while they steal only 2-3 percent points when they are in position 2–4. Hence, removing the own-brand ad would on average lead to the loss of 15 percent points of clicks.

perhaps, some firms do not know whether the results of eBay experiment apply to them. Thus, instead of reacting by shutting down their brand search advertising, these firms decide to experiment and learn brand search advertising effectiveness on their own. This leads to the second reaction we study: do firms adopt active experimentation and increasingly evaluate the ad effectiveness for themselves? The "search pause" methodology described in BNT is incredibly easy to adopt because it can be done through the bidding interface with no additional programming (more advanced methods are discussed in section 2.3). The ease of experimentation was aptly summarized in the popular press and held up as a contrast to the staggering value it offered eBay. We can identify experimentation in our data by looking for sharp changes in advertising levels—indeed we are easily able to identify the dates of eBay's experiments. In general, sharp changes could reflect active experimentation, or "natural experiments" due to budget exhaustion, churn in marketing personnel and so forth. Since the value of experimentation might not depend on the presence of competing ads, we compare the frequency of experiment-like variation in advertising levels before and after information disclosure. This produces precise estimates that reveal no significant difference, indicating, perhaps surprisingly, that firms were not moved to adopt this powerful method. A difference-in-differences approach leads to similar point estimates and the same conclusion.[3]

Although experiment-like changes in advertising did not increase, they were nonetheless quite common—72% of the 395 of the firms facing limited competition had at least one sharp change in advertising levels over the five year period. One possibility is that firms do not appear to react because they were already conducting experiments. In this case, the findings of BNT would not reflect new information, and while this runs contrary to the narrative in academic circles and the press, it is nonetheless an important possibility to investigate. We use sharp changes in the firms' advertising to estimate advertising causal effects at the firm level, restricting ourselves to data that would be available to advertisers. The resulting estimates reveal firm-level heterogeneity in the ad effectiveness and are consistent, in terms of overall magnitude and distribution, with the estimates from randomized trials on the same platform (Simonov et al., 2017).[4] Most estimates are close to zero (2-4 percentage points),

[3]We note that the advertising level changes that we describe above can reflect firms' experimentations that resulted in shutting down brand advertising, meaning that the BNT coverage has triggered some experimentation. However, we can conclude that the volume of these incremental experimentations was not high enough to stand out among the normal changes in the frequency of brand advertising.

[4]We define an "ad effectiveness" study as one that measures the causal impact of advertising, calibrated in terms of dollars, versus the cost of the media. Given that we do not observe the profitability of each customer to the focal brands, we assume that clicks on the focal brand's weblink have similar profitability and proxy for profits with the overall volume of traffic navigating to the focal brand's website.

but some firms have ad effects significantly above the mean, and others observed a more tightly estimated zero. Strikingly, the magnitude of the observed ad effect has no predictive power for advertising levels in the last six months of our data, which we use as a hold-out evaluation period. In other words, even firms that possess data to tightly estimate small ad effects show no differences from firms that had ad effects significantly greater than the mean effect. The lack of relationship between the ad effect a firm could potentially measure and future advertising levels stands in contrast to eBay's reaction to discontinue advertising on own-brand brand queries.

Putting all the pieces together reveals a nuanced picture. First, the negative information on brand search ad effectiveness led to a reduction in the propensity to advertise on own-branded queries. While this impact is highly significant statistically and economically meaningful, the majority of firms nonetheless continued with business as usual, pointing to substantial "inertia" of business practice. Second, there is no measurable impact on the propensity to conduct in-house experiments nor do the insights they could have easily derived from existing experiment-like variation in ad levels appear to impact decisions. The natural deduction is that such variation either does not represent intentional experimentation, analytics teams conducting these experiments do not make the right inference, or there is a principal-agent problem within the firm preventing the gathered evidence from guiding decisions. Such agency problems could also contribute to the failure to start actively experimenting to measure ad effectiveness despite the positive information about the value of experimentation.

Past work has theoretically explored how agency problems can arise when the incentives of the proximate decision makers diverge from those of the firm as a whole (Scharfstein and Stein, 2000). In our application, it is marketing managers who are tasked with optimizing and evaluating advertising expenditure. While search advertising experiments on own-brand queries are clearly quite valuable at the *firm level*, at the *individual level* they could produce "bad news" by revealing past mistakes or reducing the overall expenditure on digital ads (most firms have separate budgets for digital "performance" advertising vs. brand advertsing on media such as TV).[5] In contrast, a high-employee outside the marketing organization, such as the Chief Financial Officer, has a more clear incentive to learn true return on advertising

[5]Reporting incentives have been previously studied theoretically. An example is the persuasion game of Shin (1994). The analog here would be a marketing manager opting to continue to report nominal CPC and CTR, not incremental traffic induced by the ad, or to selectively report experimental results.

investment. Such an employee could read an article in an outlet like the *Harvard Business Review*, quickly fire off a few tests searches to observe what ads, if any, appear on the firm's brand queries and, if necessary, order own-brand ads to be shut down. This could be done with the "flip of a switch" and would not require changing the culture of measurement at the firm.

To the best of our knowledge, this is the first paper to quantitatively measure how firms react to scientific findings on business practice. Business practices have, in turn, been linked to the large differences in productivity per worker within narrowly defined industries that cannot be attributed to differences in capital (Bartelsman and Doms, 2000; Foster et al., 2008), such as using measurement and monitoring practices in-line with established best practices (Bloom and Van Reenen, 2010). Such best practices are often established by within-firm field experiments similar to the one we study in this paper (see Bandiera et al. (2011) for a review).[6] The large and growing body of field experiments of this sort have addressed topics such as how managerial practices can improve supply chain efficiency (Bloom et al., 2013),[7] evaluation of performance-based pay (Lazear, 2000; Shearer, 2004), manager incentives (Bandiera et al., 2007), the interaction of social preferences and incentives (Bandiera et al., 2005; Ashraf et al., 2014) and optimally running auctions (Ostrovsky and Schwarz, 2011).

Perhaps the closest related work are findings from the field of medicine that suggest doctors' choices among competing treatment options are broadly consistent with rational information processing with knowledge spillovers (Chandra and Staiger, 2007). Fiedler (2013) argues that new treatment adoption is consistent with weighing the evidence in published papers and a doctor's personal experience. In contrast, the knowledge spillovers arising from the public disclosure of the BNT findings are present but modest in size and incomplete on some dimensions. The differences between these two settings could help illuminate the root causes of the learning dynamics we observe. In medicine new techniques can be good for patients, doctors and hospitals alike. Further, doctors are highly trained to process technical information about the efficacy of new treatment options, actively contribute to the academic literature, and hospitals are themselves often institutions of higher learning. Advances in business practice, such as the one we study, are not the result of new inventions, but rather

[6] Bandiera, Barankay and Rasul (2011) note that such experiments often have a large impact on profits when the experimental findings are implemented by the firm.

[7] Particularly relevant, the authors conclude that changing incorrect beliefs about important practices is the main driver of the effect.

new approaches to solving existing problems. It is commonly argued that agency problems, such as those we highlight for advertising, are generally present within large firms because the incentives of workers and divisions within a firm often are different from those of the broader organization (*e.g.* Scharfstein and Stein (2000) and Lazear (2000)). If so, our results may point to a broader drag on economic efficiency and could be an important factor in explaining the wide variation in worker productivity in narrowly defined industries.

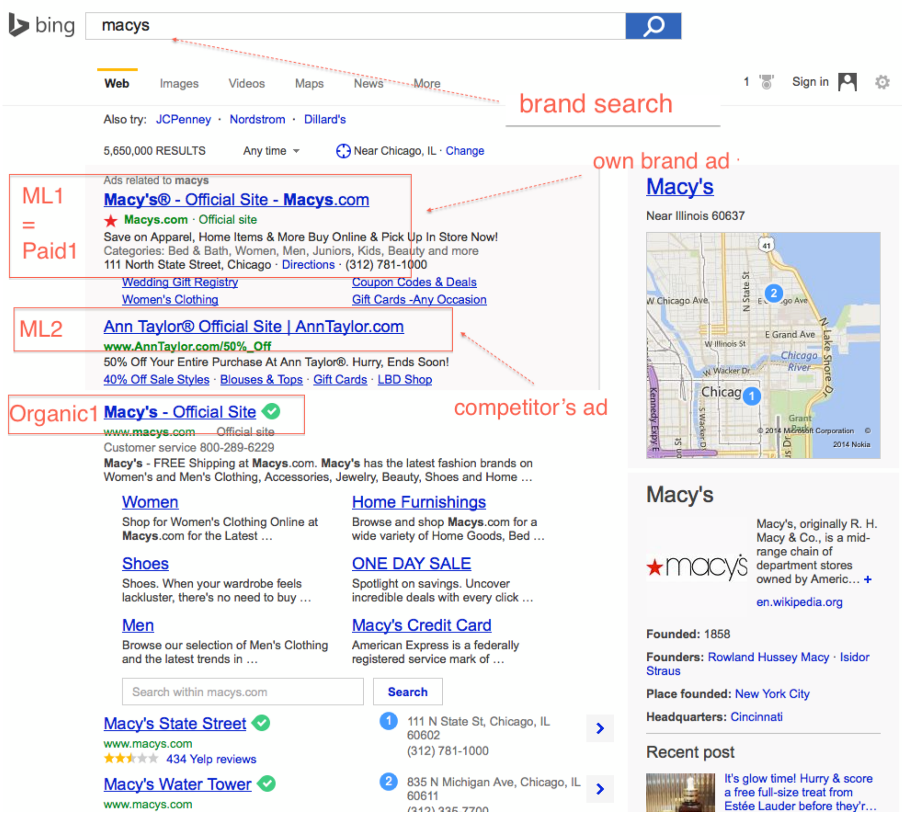
The paper proceeds as follows: Section 2 describes the economic setting and the data, Section 3 gives results for the propensity to advertise and experiment, Section 4 estimates advertising effects at the firm-level and examines the correlation of these estimates with companies' post-experimentation behavior, and Section 5 concludes with a broader discussion.

2 Economic Setting and Data Description

In sponsored search, advertisers bid on keywords to have their ad displayed at the top of search engine results. An advertiser specifies keywords that determine which queries a given ad will be entered into a real-time Generalized Second Price auction. The search engine sets parameters such as the reserve price, relevancy requirements for the ad and the maximum number of allowable ads that appear above the organic listings (currently four for Google and Bing). "Brand search" advertising occurs when an advertiser bids on keywords that include a brand name. Figure 1 gives an example for the department store Macy's advertising on the keyword "macys." Macy's paid link occupies the top position on the page and a competing advertisement occupies the second slot. Importantly, Macy's also appears in the first organic position on the page. All clicks on the organic results are free.

Brand search is an attractive domain to study the effect of the information disclosure for a few reasons. First, BNT report that the potential cost savings are large—eBay management decided to discontinue expenditure of over fifty million dollars per year. Second, the inference problem is challenging because "nominal" metrics of ad performance, such as click-through rate (CTR) and cost per click (CPC), tend to overstate the actual advertising effect (Lewis et al., 2015), creating a large wedge between the reported price (CPC), which is typically very low for the focal brand low due to the high relevance, and real "cost per incremental

Figure 1: Example of Macy's advertising on own brand keyword Macy's.



click."[8] Since these real metrics are not directly observable, firms must conduct experiments to estimate them, just as eBay did. Third, published evidence on the effectiveness on own brand keyword was limited until the release of the eBay study.[9] Finally, at the time BNT was released, firms had heterogeneous behavior with respect to own-brand search: roughly half of reasonably large firms that otherwise used search advertising chose to advertise on their own keywords, and the other half chose not to (Simonov et al., 2017).

2.1 Blake, Nosko and Tadelis (2015) and Media Coverage

Blake, Nosko and Tadelis (2015) reported the results of a large field experiment in which eBay shut down their brand keyword advertising for several weeks of spring 2012. The results revealed search ads on queries containing "eBay" had almost no effect on traffic to eBay's website. The paper was released as a working paper at the beginning of March 2013 and was quickly picked up by the business press. A detailed write-up appeared in *Harvard Business Review* on March 11, 2013, with the provocative headline shown in Figure 2 and was quickly followed by the coverage in the BBC (March 13), *Business Insider* (March 14), various smaller outlets and leading search engine blogs. *The Atlantic* and *The Economist* ran feature stories in April and July 2013 respectively. For the purposes of the information disclosure data, we use March 11, 2013.[10] While the coverage often started with an attention grabbing headline, the articles generally gave a faithful description of the study and reasonable recommendations for business practice. BNT had a large impact in academic circles as well, securing publication in a top journal, *Econometrica*, and was one of five finalists for the 2015 Gary L. Lilien ISMS-MSI Practice Award, which recognizes academic papers in the marketing discipline that have the biggest impact on marketing practice.

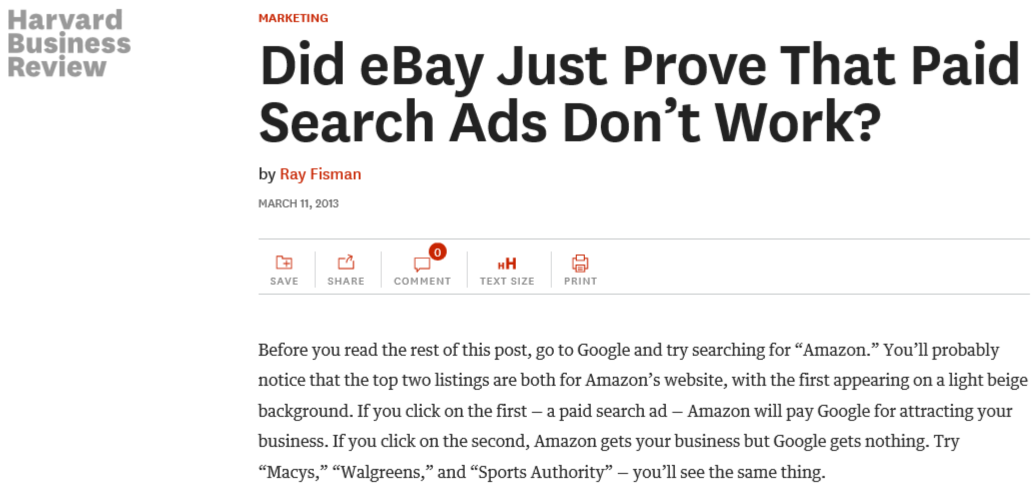
BNT ran their experiment on Bing and Google, meaning the "traces of experimentation"

[8] The pricing rule of the GSP rewards high CTR ads with low CPCs, which makes sense since the opportunity cost of the search engine is impressions, see Edelman et al. (2007) for more details. Simonov et al. (2017) show that in the absence of competitors ad, own brand advertisement increases the probability to get a click by 0.014 while a naive measure would estimate a 0.4 probability increase.

[9] A quick web search reveals practitioners' guides that warn of the problem of click crowd-out and others that recommend advertising on own-brand keywords. On the academic side, until March 2013, there was only one paper examining the interdependence of paid and organic traffic (Yang and Ghose, 2010), which used observational data to conclude, counter-intuitively, that clicks on paid links actually increased clicks on organic links (crowd-in).

[10] Titles of the media coverage are listed in Appendix 6.1.

Figure 2: Harvard Business Review article on the eBay search experiment. The story was published March 11, 2013 (hbr.org/2013/03/did-ebay-just-prove-that-paid).



Harvard Business Review

MARKETING

Did eBay Just Prove That Paid Search Ads Don't Work?

by Ray Fisman

MARCH 11, 2013

SAVE SHARE 0 COMMENT TEXT SIZE PRINT

Before you read the rest of this post, go to Google and try searching for "Amazon." You'll probably notice that the top two listings are both for Amazon's website, with the first appearing on a light beige background. If you click on the first – a paid search ad – Amazon will pay Google for attracting your business. If you click on the second, Amazon gets your business but Google gets nothing. Try "Macys," "Walgreens," and "Sports Authority" – you'll see the same thing.

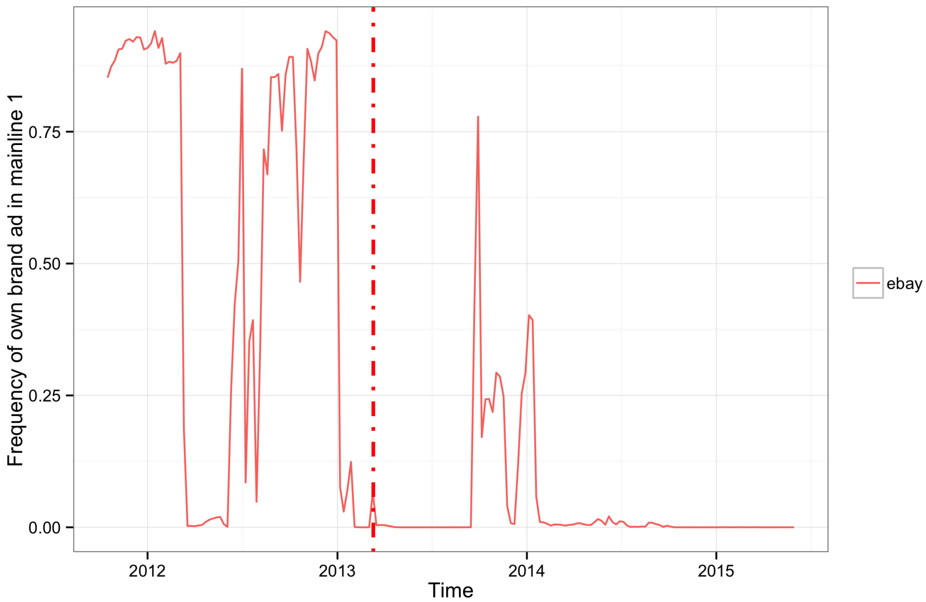
should be observable in our data and indeed they are. Figure 3 shows the frequency which eBay advertised on their own brand keywords over time.[11] Since the authors provided a detailed description of the experiment and the subsequent reaction of the company, this figure does not reveal anything new. Prior to Q2 2012 eBay regularly advertised on their keywords. The experiment is visible just at the time period reported by BNT, where we see own-brand advertising paused, after which there was a return to business as usual before the ads were discontinued in 2013, outside a brief period of what appears to be a follow-up experiment. This check validates that our data can capture experimentation behavior and is entirely consistent with the description in BNT.

2.2 Data Description

We start with 87,000 brand names of online retailers taken from the Open Directory Project. We aggregate historical search logs from individual event level (e.g. a page view) to the the daily level from October 07, 2011, to May 31, 2015, which both anonymizes the data and reduces it to a practical size. The resulting sample consists of 85,725 brands searched at

[11] In Figure 3 and throughout the paper, we restrict our attention to keywords that contain only the brand's name or brand's website (e.g. "eBay" or "ebay.com"), not more broad brand-related searches (e.g. "eBay shoes"), since the results of BNT have the strongest implications for the former set of queries.

Figure 3: eBay's own-brand keyword advertising frequency over time.



least once during this period. We further limit to firms that (1) are searched consistently, (2) are in the top organic position on their own-brand queries, (3) advertise on their brand keywords at least some of the time and (4) "want the traffic" (that is, the brand is not generally sold through a reseller). Specifically, we reduce the sample by keeping companies which (1) are searched for at least 20 times on each day in our sample (6,258 companies), (2) are in organic search position 1 more than 90% of the times (1,861 companies), (3) advertise in top advertising position more than 90% of the time on at least one day in the sample (1,234 companies) and (4) get at least a 50% combined CTR from the ad and organic links (1148 companies), meaning that consumers want to navigate to this firm's website and not to potential resellers. These restrictions leave us with firms for which brand-search is relevant (those that get meaningful traffic), high-quality firms (a firm should be the top link on its trademarked terms), advertise on Bing and are not firms that sell through re-sellers.

Table 1 gives summary statistics. An average brand gets more than three million searches over the time period we study. There is more than one mainline ad on average, with a standard deviation of 0.9. This implies our data has a lot of variance in the number of ads shown in the advertising slots above the sponsored positions. We further see that there is variation in both percents of the own brand ad in mainline 1 occasions and competitor ads

Table 1: Data Summary.

Variable	Mean	Median	Standard deviation
Exposures	3,262,494	575,711	16,969,288
# of mainline ads	1.67	1.41	0.9
% of times own ad in mainline 1	71.35	81.62	26.2
% of times competitor's ad in mainline 1	12.37	6.93	14.82
% of times competitor's ad in mainline 2	38.14	30.9	27.53
% of times competitor's ad in mainline 3	25.47	13.39	27.4
% of times competitor's ad in mainline 4	16.77	5.16	23.67
% of search with navigation to own brand website	74.60	77.08	8.85
- through organic position 1	46.84	46.11	14.6
- through mainline 1 ad	27.76	29.42	14.3
Number of observations		1,525,692	
Number of days		1,329	
Number of companies		1,148	

Means are unweighted.

in mainline 2-4 occasions. The percent of search occasions leading to an own-brand click (paid or organic) is around 75 percent, which is consistent with previous results in Simonov et al. (2017).

2.3 How to Purchase Search Advertising

Ads on search engines can be purchased in a variety of ways. The simplest option is to use the "portal" provided by the search engine. Here one can choose keywords, bids, set budgets and view statistics on current ad listings, such as average position, CPC, CTR, etc. The key choice variables can also be set "programatically" via software that hooks into portal APIs. A firm can choose to either assign the task of keyword selection, budget allocation and bidding to specialist employees in the marketing department (who use either the portal or software), or hire an advertising agency that specializes in "search engine marketing" to handle most of the day-to-day operations. In this case, an employee within the firm is typically assigned the responsibility of sharing data with the agency and managing the relationship. So in terms of who is actually doing the bidding, by-and-large it is either an employee in the digital division (if such a division exists) of the marketing department or an ad agency that gives feedback on advertising performance through to their counterpart at the firm.

The first step to conduct search advertising experiments is to define a set of keywords (line item "bids") to include in study. For example, in our setting this would be all search terms containing the firm's brand name. Next, there are three, broadly speaking, experimental designs a firm can use to test the effectiveness of search advertising: 1) temporal variation, all bids for the selected keywords are turned off for a period of time, or on alternating days or hours within the day 2) keyword hold-out, a randomly select a set of keywords from the original set to continue business as usual bidding 3) regional hold-outs or "sister cities," search ads can be geographically targeted, so a collection of holdout cities. All three methods induce fewer ad purchases for the specified keyword set, so are observable by the platform.

3 Estimation and Results

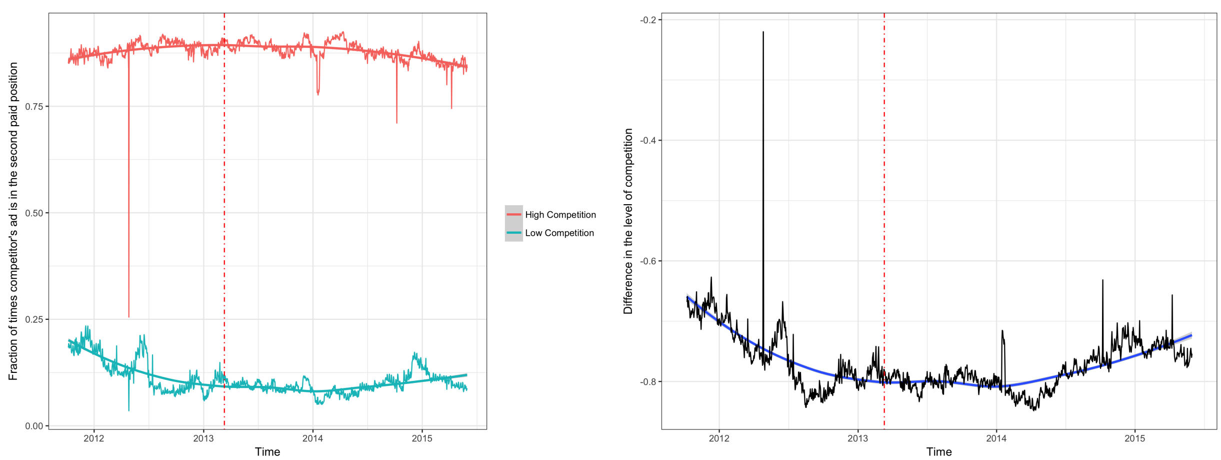
We start by studying the impact this publication had on firms' propensity to engage in own-brand advertising. Our identification comes from the observation that companies which face the same level of competition as eBay—having no competitors advertising on their keyword—should be affected by this information, while companies which generally face competing ads are not directly affected. The reason is simple: when a firm faces competing ads, if it kills its own-brand ad, a competitor will supplant it at the top of the page. Simonov et al. (2017) show that in such cases, competitors can siphon off 20% of clicks on average, an order of magnitude higher than the average causal impact of own-brand ads without competitors present. This implies that in this case removing own brand ads would lead to losses of clicks far higher than that reported by BNT, who are careful to highlight the critical importance of competing ads (or the lack thereof in their case).

Figure 11 in Appendix 6.2 shows the distribution of firms over the frequency of facing a competitor in slot 2. There are companies for which competitors rarely advertise in Mainline 2 and companies with competitors frequently advertising in Mainline 2. We define companies which face a competitor's ad in Mainline 2 less than 20% of the time as companies with a low competition level, and companies which face a competitor in Mainline 2 more than 80% of the time as companies with high competition. The resulting group with the low level of advertising competition (which we refer to as “treatment”) contains 395 companies, and the group with high level of advertising competition (“control” or “untreated”) contains 145 companies.[12]

3.1 Verifying Diff-in-diff Identifying Assumptions

We start with examining the validity the identifying assumptions for our “treatment” and “control” groups. The design is valid provided there was not a confounding factor around the time of information disclosure that differentially (by group) affected the propensity to conduct own-brand keyword advertising. Since we define groups using the level of competition, we examine if the extent of competition changed around the time of disclosure and the time trends in each group leading up to disclosure.

Figure 4: Competition level for “treatment” and “control” companies: frequency of competitor ad in Mainline 2.



(a) Levels (b) Difference

Fraction of times competitor ad is shown in Mainline 2.

Figure 4 plots the frequency of a competitor ad in Mainline 2 for groups of companies with high and low competition levels. We fit a local polynomial[13] to assess the degree of change, which is given in panel (a). The frequency of competitors advertising in Mainline

[12] Throughout the paper, we use the “level of advertising” term to refer to the frequency of ad appearance in the top paid positions and not to the amount of dollars spent on advertising.

[13] LOESS of second degree.

2 in "treatment" case was decreasing in the middle of 2012 but is stable at the time of publication, made clear in panel (b), which models just the differences in the propensity of groups to face competition. Further, the absolute changes over time in both cases are minimal.

The parallel time-trend assumption is examined in Figure 5, with the difference in the series given in the panel (b). Examining the period before information disclosure (the region left of the vertical line), the time trends are both increasing at nearly the same rate. Note the level differences are expected: firms facing competition are more likely to engage in own-brand advertising as the returns are higher in this case. Finally, the lack of a meaningful pre-period time trend is shown clearly in panel (b), which plots the differences and reveals a slight downward trend, the impact of which is entirely swamped by the treatment effect we find later.

Overall the identifying assumptions for the difference-in-differences estimation hold nearly exactly, and where they do not, the small divergence is not large enough to generate significant treatment effects.

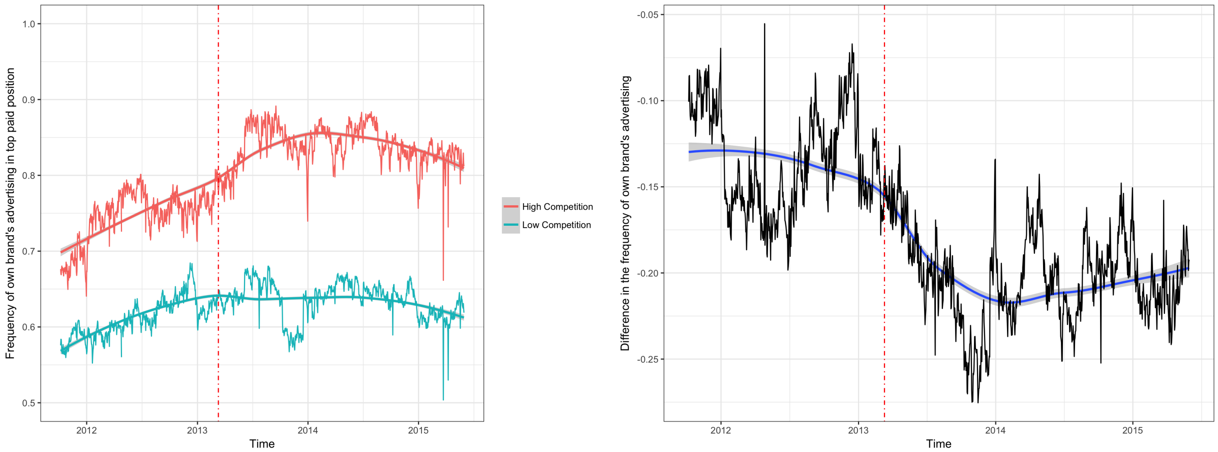
3.2 Effect on Advertising Levels

Figure 5 plots the frequency of own-brand advertising in slot 1 for companies with low competition (blue) and companies with a lot of competition (red). The frequency of advertising in the top slot increases in both groups before the information is revealed, but the low level of competition group flattens out after publication of the BNT article. The plot on the right presents the difference between groups with low and high levels of competition and non-parametrically illustrates a significant treatment effect that takes roughly one year to fully take hold.

We use a difference-in-differences estimator to provide formal tests of this effect, incorporating various controls for seasonality and brand fixed effects. The results are presented in Table 2. The estimated impact is 0.067 percent points (averaged over the entire post-period), with most reactions happening within the 120 days window after the information period.[14]

[14] Appendix 6.3 presents the estimates for different specifications of the time window around the information dissemination event as well as placebo tests.

Figure 5: Own advertisement level for "treatment" and "control" companies: average level.



(a) Levels

(b) Difference

Fraction of times own brand ad is shown in Mainline 1 for companies facing and not facing competitors in Mainlines 2-4. Daily data.

Table 2: Difference-in-differences regression on level of advertisement.

	(1)	(2)	(3)	(4)
Treatment	-0.1351	-0.1351	–	–
	(0.0281)	(0.0282)		
After	0.0926	–	0.0926	–
	(0.0188)		(0.0188)	
Treatment * After	-0.0679	-0.0679	-0.0678	-0.0679
	(0.0237)	(0.0237)	(0.0237)	(0.0237)
Focal Brand FE	N	N	Y	Y
Day FE	N	Y	N	Y

Robust standard errors clustered on focal brand level.
All coefficients presented are significant at 1% level.

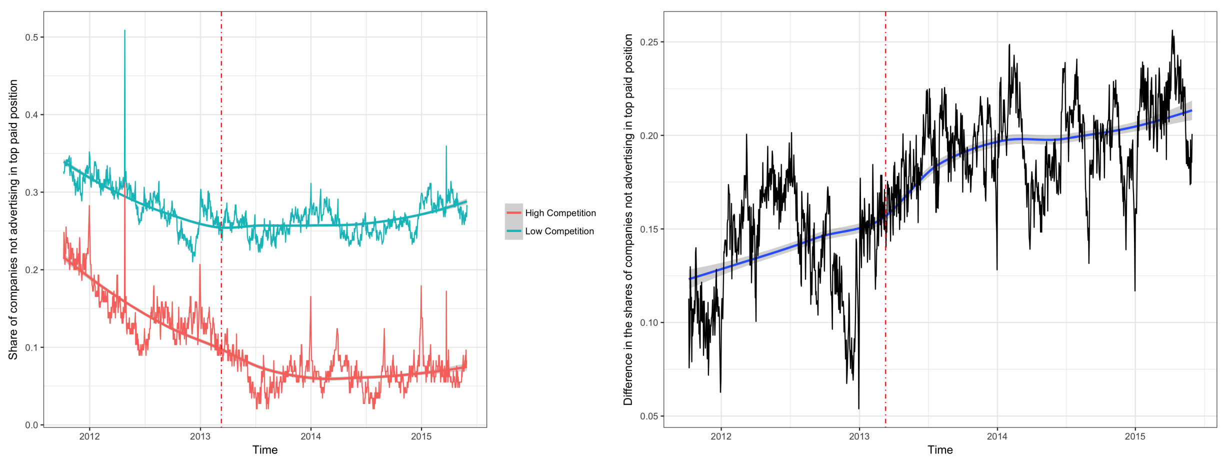
This estimate is significant beyond the 0.01 level and is very stable across specifications. Given that average level of advertising for companies not facing competition is 0.626, this decrease corresponds to 10.8% change in advertising levels.[15]

This change could come from two qualitatively different reactions, which at the extremes

[15]Figure 14 in Appendix 6.4 confirms these results by presenting the non-parametric relationship between a change in the focal brand's advertising level after the BNT disseminations and the level of competition for all firms in the sample.

are given by: (1) all companies reduce their advertising level by 11%, (2) 11% of companies completely turn off their advertising. Indeed, this distinction offers a robustness check for the results. If decreases in the level of advertisement are due to the effect of information, then we would expect the latter pattern: the ads are simply turned off, just as eBay did. On the contrary, if pattern (1) is observed, then it is difficult to reconcile as a reaction to the news in BNT.

Figure 6: Own advertisement level for "treatment" and "control" companies: fraction of companies advertising less than 20% of the time.



(a) Levels

(b) Difference

The fraction of companies with own brand ad in Mainline 1 < 20% of the time. Daily data.

We investigate these possibilities by measuring the number of occasions that companies advertise on their own query on less than 20% of searches in the week (advertising "turned off") in both groups. If companies indeed stop advertising, we should find that there are more companies without competition that move to the no-advertising state after the publication compared to the companies which face competition. Figure 6 presents levels and difference of the share of companies not advertising on their own keywords. The share of companies not advertising on their own keywords was decreasing in both groups before the publication. After the publication, this quantity stabilizes among the companies which do not face competition but it keeps declining among the companies which do face competition. The difference between treated and untreated firms is given in the right panel of Figure 6, which shows the clear relative uptick in not advertising for the treated firms.

To formally test the significance of this visual evidence we use a difference-in-differences

Table 3: DiD regression on occasions of not advertising.

	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.829	–	0.9413	–	0.8389	–
	(0.1871)		(0.1897)		(0.1905)	
After	-0.9077	–	-0.9092	–	-0.9082	–
	(0.1879)		(0.1878)		(0.1876)	
Treatment * After	0.7755	1.0768	1.138	1.4095	0.8379	1.1942
	(0.2038)	(0.3165)	(0.4581)	(0.7112)	(0.2027)	(0.3205)
log(US)			0.0956	–		
			(0.0433)			
Tr. * Aft. * log(US)			-0.0433	-0.04		
			(0.0499)	(0.078)		
C. Response					-0.14837	–
					(0.3086)	
Tr. * Aft. * C. Response					-0.9153	-1.4393
					(0.3974)	(0.8212)
Focal Brand FE	N	Y	N	Y	N	Y
Day FE	N	Y	N	Y	N	Y

Robust standard errors clustered on focal brand level.
All coefficients presented are significant at 1% level.
Link function is logit.

logistic regression with "not advertising" (=1) as the binary outcome variable. The results are presented in Table 3, columns (1-2). The effect for treated firms is significant beyond on the 0.01 level, and the corresponding average marginal effect of information on the probability of not advertising is 10.4%.[16] This indicates that nearly all of the advertising level effect we previously observed is driven by companies discontinuing own-brand ads, consistent with an informational story.

While on average firms react to new information and stop advertising, the number of firms reaction is rather modest given the strength of the information. One natural concern a firm could have is that there is heterogeneity in ad effectiveness across companies and that eBay's results might not apply to every firm. Indeed, Simonov et al. (2017) show that the effect of a company's ad on its own brand query is different across companies; in particular,

[16] Based on specification (2). For robustness, we also check another measure of the level of advertisement. Figure 15 in Appendix 6.5 plots the fraction of companies with own brand ad in Mainline 1 > 90% of the time for "treatment" and "control" group. Results are the same.

the effect is larger for companies with less brand capital. Perhaps firms are aware of these differences in ad effectiveness, so firms that are less similar to eBay are more reluctant to follow eBay's example.

To test whether our estimates of firms' advertising level reaction are driven primarily by companies similar to eBay, we interact the difference-in-differences estimator with brand prominence. Following Simonov et al. (2017), we proxy for the focal brand's prominence with the U.S. rating of websites' popularity from Alexa.com. Columns (3-4) in Table 3 present the estimates. We find no evidence that more prominent brands tend to react less than an average brand.

Another anecdotal reason for firms to advertise on own brand traffic is the belief that it deters competitors from entering and advertising. Such competitive reactions are possible due to the ability of firms to set up bidding systems that condition on the recent presence of other bidders.[17] At the same time, it is not clear why such behavior is optimal for competitors, since the presence of the focal brand's ad does not affect their profit margins (the focal brand almost always wins the auction due to its high relevance and expected CTR).

We first examine whether there is evidence of focal brands' ads serving as deterrents. For this, we regress an indicator variable of competitors' advertising in mainline 1 or 2 on an indicator variable of the focal brand's advertising.[18] There is a strong negative correlation in the presence of competitors' and the focal brand's ad. Figure 16 in Appendix 6.6 shows the histogram of changes in the probability of competitors' advertising when the focal brand stops advertising. On average, competitors are 8.4 percentage points more likely to advertise when the focal brands do not. Competitive entry is prominent for a minority of firms, with 31% of focal brands experiencing a more than 10 percentage points increase in the probability of competitive advertising when they stop their brand advertising and 8.4% of companies experiencing a more than 50 percentage points increase.

[17] Indeed, such behavior was widely observed in the early days of search advertising when first-price auctions were used (Edelman and Ostrovsky, 2007).

[18] We regress I(Competitor in ML1 $>$ 50% or Competitor in ML2 $>$ 50%) on I(Focal Brand in ML1 $<$ 10%), standard errors clustered at the focal brand level, focal brand and date fixed effects included. The design of the dependent variable ensures that the absence of the focal brand's ad does not affect the measure of competitors' advertising mechanically. Both 10% and 50% thresholds are chosen arbitrarily; the results are robust to the deviations from these thresholds.

We then test whether the firms that face a greater threat of competitors' entry are less likely to react to the BNT coverage. For this, we interact the difference-in-differences estimator with the focal-brand-specific change in the probability of competitors' entry.[19] Columns (5-6) in Table 3 present the estimates. Focal brands that face the threat of competitive entry are less likely to respond to BNT, with the effect significant in specification (5) and marginally significant (p-value = 0.079) in the specification (6). The magnitude of the estimates implies that if the competitors will definitely enter when the focal brands stops advertising (C. Response = 1), the focal brand does not react to the BNT coverage.[20]

3.3 Effect on Experimentation Levels

In the previous section we have examined the firms' reaction to the BNT information in terms of the advertising levels and have explored the heterogeneity of this reaction. However, in practice the firms might not know their "type" and might hesitate about whether or not they should follow eBay's example. Given the ease of implementing BNT's experimental method—the baseline version can be done by just pausing the ad through the bidding interface with no programming required—the rational response for these firms would then be to copy the protocol and learn one's own ad effect. We now examine if firms indeed reacted in this fashion.

As shown in Figure 3, intentional experimentation is characterized by large, sharp changes in advertising levels. These changes, however, do not necessarily represent experiments, as they could be due to budget shortfalls, change in personnel that handles the ad buying and so forth. Accordingly, this "experiment-like" variation is a necessary but not sufficient condition for intentional experimentation. We start by looking at the amount of experiment-

[19] As before, we define this competitor's response as

$$\text{C. Response} = \Pr(\text{C. Advertise}|\text{Focal Brand in ML1} < 10\%) - \Pr(\text{C. Advertise}|\text{Focal Brand in ML1} > 10\%),$$

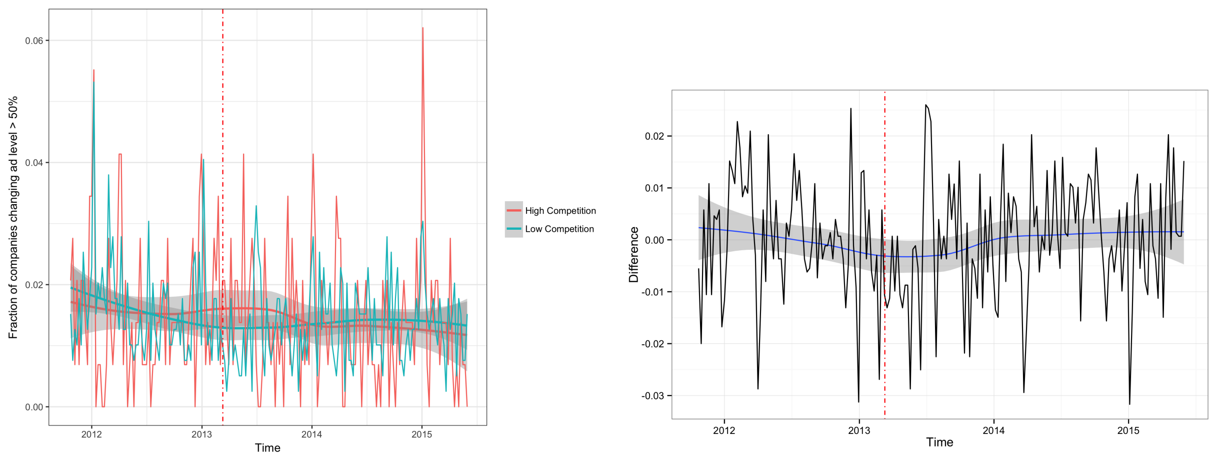
where

$$\text{C. Advertise} = \Pr(\text{Competitor in ML1} > 50\% \text{ or Competitor in ML2} > 50\%).$$

[20] We note that our estimates of competitive entry are based on correlations. To confirm that these competitive reactions are causal, we examine changes in the competition level right before and after the focal brand's change in advertising, which are the discontinuities used in Section 4.1. We confirm that competitors start to advertise right after the focal brand stops. Such fast reaction is consistent with the algorithmic response story that we have discussed above.

like changes in advertising before and after the experiment for companies with high and low levels of advertising competition. If companies similar to eBay started to experiment more, we expect the total amount of experimentation to change. In this subsection we focus on weekly level data to avoid variation in the advertising level due to weekdays and weekends. We define experiment-like changes as a 50 percentage point difference in advertising levels week-on-week (our results are not sensitive to the precise definition).

Figure 7: Fraction of companies changing their advertisement level by more than 50%.



(a) Levels

(b) Difference

Weekly data.

Figure 7 plots the weekly propensity of experiment-like variation over time. The average propensity is 2% per week, and the time trends reveal that there is no detectable increase in the experiment-like changes after BNT was released. A difference-in-differences specification confirms there is no significant difference for the treatment group relative to the untreated group and a simple before-and-after estimator reveals no significant differences for either group.

Although we do not observe an uptick in the experiment-like behavior, large changes in advertising week-on-week are not entirely uncommon (2% of weeks, or about once-a-year per firm). Out of 395 companies that do not face competition, more than 71% have changed their advertisement level by more than 50% at least once in the observed period. One reason a firm may adjust their advertisement levels is the existence of demand shocks (*e.g.* holiday shopping period or product releases). If expected demand changes affect the level of advertising, we should find a strong correlation between the change in the number of total query exposures in a given week and change in the advertising rates. The correlation between percent change in exposures and change in advertising rates that we find explains less than 0.1% of variation of changes of advertising levels. This allows us to conclude that demand factors are not the primary reason for the changes in advertisement levels that we observe.[21]

The results here indicate that there was no experimentation uptick but that a majority of firms nonetheless had large changes in advertising levels suitable for estimating causal effects. This either reflects widespread experimentation prior to BNT or natural experiments due to the supply-side mechanisms (*e.g.* running of out budget). Although the narrative of coverage of BNT and the impact it has had in academic circles was around the innovative nature of the experiment and findings, it's possible, albeit unlikely, that most firms had been conducting these experiments all along. This would explain why a small percentage of firms reacted in terms of advertising levels and the lack of a reaction in terms of experimentation levels. In the next section we examine if these sharp changes in the advertising levels were indeed active experimentation by computing the implied advertising effectiveness and if firms changed their advertising levels in response.

4 Firm-level Estimates and Reactions

We start by computing firm-level advertising effects for companies which have the requisite variation in ad levels. We then examine their behavior after "experimentation-like" variation in light of these estimates. If the experiment-like variation we observe indeed represents intentional experiments, rational firms that find large effects should decide to keep advertising, and those that find small effects (the majority of firms) should stop advertising. Examining this relationship lends insight into whether the variation we observe represents intentional, well-run experiments or is due to other factors such as budget constraints.

[21] Recall that all results for this section are based on weekly data. If we use daily data, results would be different in this case: we find that demand changes explain around 2% of all advertising level changes in daily data. This is explained by some firms treating weekdays to weekends (lower query volume) differently.

4.1 Firm-level Advertising Effects

Our dependent measure is total traffic to the brand's website from the sponsored and organic links.[22] On average, 80% of searches result in a click to the own-brand website. Figure 8 presents the histogram of probabilities of getting to the focal brand's website for companies that do not face competition, plotted separately based on whether the firm advertised on more than 90% of the occasions in a given week and when they advertise less than 10%. The distribution is shifted to the right during advertising spells, and this difference is significant (Kolmogorov-Smirnov, p-value = 2.58e-12). During advertising spells, the average probability of a click is 81.5%, and on occasions without advertising it is 78%, implying a 3.5 percent point difference between advertising and no advertising cases. This simple comparison reveals a correlation between advertising and the probability to get clicked, which could be due to the causal effect of the ad or endogenous timing of advertising.

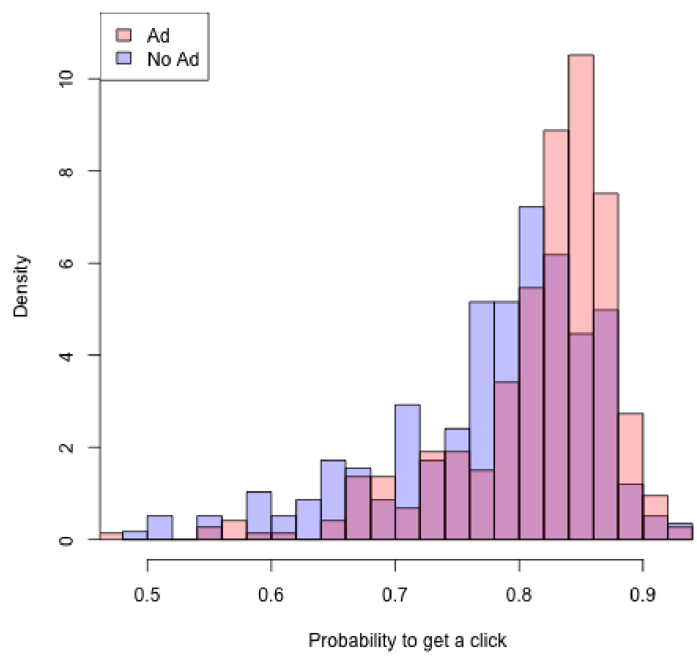
To help mitigate this potential bias, we impose additional restrictions to better identify ad effects with experiment-like variation. To do so, we use only weeks on the border of the large changes in advertising, before and after ads are turned on/off, to reduce any co-varying factors which change over time. We emphasize that this is not as good as fully-randomized experiment as there can be different demand shocks in two subsequent weeks, but we will be able to compare these estimates to the ground truth from fully randomized trials on the same platform. To control for differences in levels across the "experiments," we use changes in advertising levels and probability of click. This gives us 1081 experimentation occasions for 282 companies. Table 4 gives the results. On average, advertising increases the probability of getting a click by 2.9 percent points, similar to the above estimate. Simonov et al. (2017) estimate the causal effects of own-brand ads on Bing using full randomized experiments and find the average effect of advertising on the probability to be clicked of 1.4-2 percent points depending on firm size, indicating that while these estimates are certainly the right order of magnitude, they are potentially biased upwards by 50-60%.[23]

We examine the heterogeneity of advertisement effects across companies by running the

[22] While firms might evaluate ad effectiveness in terms of the changes in profit, we proxy the ad effectiveness with the total traffic because (1) we do not observe profit changes for the studied firms and (2) incremental profits are driven by incremental traffic (e.g. based on BNT and a case of eBay incremental profits and traffic are highly correlated).

[23] We know there is heterogeneity across companies from Simonov et al. (2017), and thus part of the difference could be due to firm composition.

Figure 8: Histogram of probability to get a click on own brand query across companies, by level of advertising.



Weekly data.

Table 4: Effect of advertising on probability to get a click.

	(1)	(2)
Advertising Level	0.0295	0.0297
	(0.0022)	(0.0028)
Focal Brand FE	N	Y

Robust standard errors clustered on focal brand level.
All coefficients presented are significant at 1% level.

regressions at the firm level. Since the majority of companies experiment only once or twice, the power of this estimation is limited.[24] Figure 9 panel (a) presents the distribution of company-specific advertising effects. The distribution is skewed to the right, and a battery of statistical tests reject that it is not Gaussian.[25] In addition, in panel (b) we examine the histogram of p-values of the estimates of company-specific advertising effects deviations from the mean and easily reject that it is uniform.

Overall, the estimates that firms could form using the experiment-like variation are close to what we have measured using full randomized trials and exhibit heterogeneity consistent with randomized trials. We thus conclude that, while being noisy and potentially biased, the estimates contain valuable information about ad effectiveness in this setting.

4.2 Post-Experiment Behavior

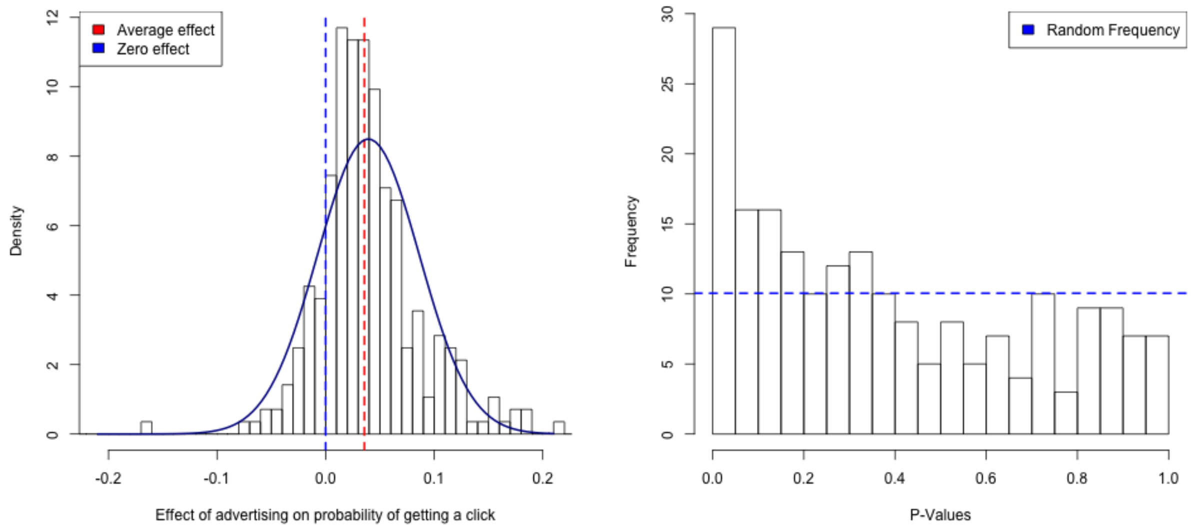
If companies are rationally using inferences from the experiments, then they should be more likely to stop advertising when advertisement effect is small and to keep advertising when the effect is large. On average, the experimental estimates are small relative to the putative ad effects given by nominal metrics (e.g. CTR or CPC), so we might also expect that most firms with experiment-like variation to discontinue advertising as eBay did.

We separate the last five months in our data and use them as the post-experimentation period and use estimates from experiment-like variation in prior periods to predict advertising levels in the holdout period. Figure 10 presents the relationship of estimated effects (y-axis)

[24] Figure 17 in Appendix 6.7 presents the histogram of number of experiments per company

[25] To conduct these tests, we normalize the distribution by mean and standard deviation, and perform a series of test: Shapiro-Wilk, Jarque-Bera, D'Agostino, Lilliefors, etc.

Figure 9: Histograms of advertising effects across firms and p-values of the estimates of firm-specific deviations from the mean.



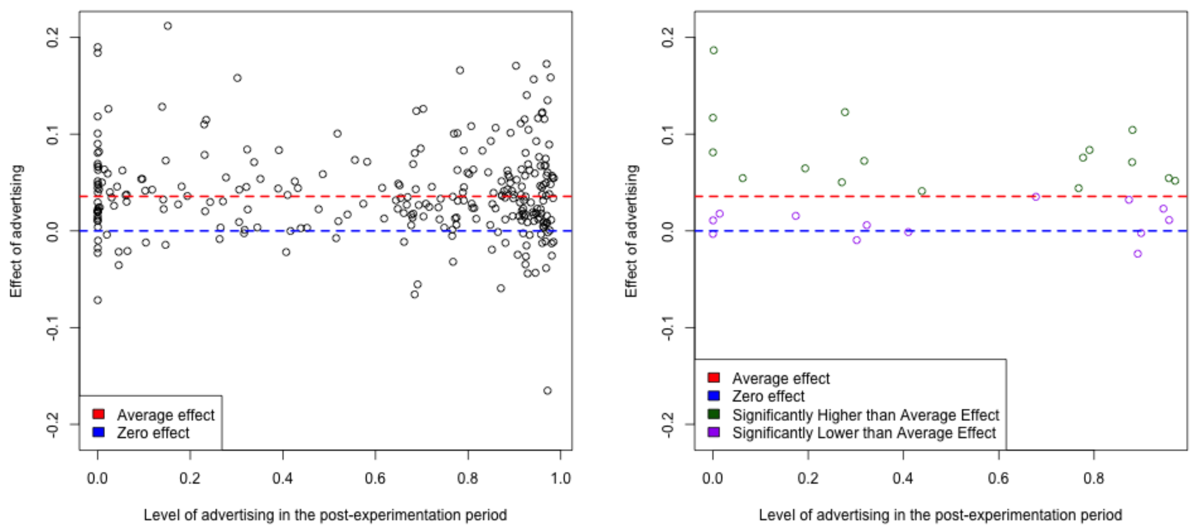
(a) Effect Distribution

(b) p-value Distribution

and their decisions to advertise in the post-experimentation period. Panel (a) gives all estimates; panel (b) restricts to those firms with ad effects that are statistically significantly different from the group mean. Companies which do not advertise in the post-experimentation period are on the left and companies which advertise are on the right. There is no detectable correlation with estimated advertising effect and the propensity to advertise, which can be seen visually and is confirmed with formal tests. This is true both for the whole sample and when restricted to firms with ad effects that differ significantly from the mean effect. Quite strikingly, the purple points in panel (b) represent firms that have a tight estimate of a near zero ad effect, and yet these firms are no less likely to advertise in the post-experimentation period than the firms with a statistically significant positive effect.

It is a reasonable deduction, then, that firms are not acting on the inferences that can be made from the experiment-like variation we observe, making it unlikely they are true experiments. As a check of this explanation, we re-run the analysis using only cases where companies did not change the advertising level for at least three weeks before and after the sharp change. Example of such behavior is presented in Appendix 6.8 (Figure 18). This behavior is more consistent with eBay's experimentation (Figure 3). We find similar results

Figure 10: Relationship between estimated effects for companies and their decision in the post-experimentation period.



(a) All estimates

(b) Estimates that are statistically significantly different from the group mean

as above in this sample of 97 companies.[26]

We cannot rule out the possibility that the proximate decision makers are conducting the right analyses, but the results are not faithfully communicated to upper management. This could be because there is an incentive to increase paid traffic generally (to increase digital ad budgets, for instance). For example, if aggregate reports feature the global average of CPC presented as "cost per visitor," then including low CPC paid traffic from own-brand queries decrease the global average. This may make search advertising look more attractive when compared to other advertising media, despite that the implicit reasoning suffers from the average-marginal fallacy. An example of this type of report is given in Table 5, which is calibrated with representive figures from Simonov et al. (2017). It is clear that by computing average CPC by pooling brand and non-brand keywords, the overall average can be greatly reduced, which may make the search engine channel appear more attractive overall. This

[26]Figure 19 in Appendix 6.9 shows the relationship between estimates of advertising effect of probability to get a click and post-experimentation decision. Also, Figure 18 corresponds to the company for which advertising effect is estimated to be negative. As we can see, the company keeps advertising in post-experimentation period.

type of strategic reporting can be seen as a classical principal agent problem.

Table 5: Calibrated example of impact of pooling reports.

	Own-brand keywords	Non-brand keywords
Cost-per-click	\$0.06	\$0.80
Ad Click-through-rate	45%	6%
Avg. CPC:	w/ brand traffic	w/o brand traffic
· 5% brand traffic	\$0.59	\$0.80
· 10% brand traffic	\$0.46	\$0.80
· 20% brand traffic	\$0.32	\$0.80
· 50% brand traffic	\$0.15	\$0.80

5 Discussion and Conclusion

Although learning by firms is central to many aspects of economic efficiency, quantitative evidence on when and how they learn and why they do not in some cases is thin. One reason for this shortage is the steep data requirements for a convincing, broad analysis. Here we have detailed data on firms' choices and the fortuitous timing of a particularly impactful academic paper, which received as much attention in the popular press as an economics paper can ever hope to, containing specific and actionable information relevant to these choices. Based on these factors, we might have expected a strong response by firms, yet our results reveal a general sluggishness to react. While we do find a significant propensity to reduce own-brand advertising for firms directly impacted by the information, only a minority of firms did so. In our discussions and presentations of this paper some commentators find our effect estimates surprisingly small, while others find the fact that we saw any effect at all surprising, especially in light of the competitors' entry threat that a fraction of firms face. We take this as evidence that it is quite uncertain how firms react to advances in business practices published in academic journals and disseminated in the business press.

We find no evidence that firms adopted the powerful (and easy) method of experimentation the paper advocated, nor do they appear to respond to inferences they could make using sharp changes in historical advertising levels (these are either natural or intentional experiments). The fact that we find a response in terms of ad levels but not experimentation points to the difficulty in moving beliefs about causal inference methods more broadly.

There are two related, and not mutually exclusive, mechanisms that could explain our findings. The first is that the relevant personnel are not paying much attention to business journals, business articles in outlets like the *Economist*, search engine management blogs and other sources of information to help optimize choices. The second is that risk aversion, the risk of punishment for past behavior and related principal-agent problems hinder the evolution of best practices at the firm. Although the policy implications to address these two mechanisms differ, they are both a product of incomplete incentives, and it is unclear which is "worse." If the relevant decision makers did not come across actionable information that received attention in the popular press, in addition to the buzz in marketing circles and technical blogs, then this is concerning since most advances do not receive near this much attention nor are their insights as straightforward to implement. On the other hand, if employees do in fact have up-to-date information but are not incentivized to act on it, this may speak to deeper problems within the firm.

It is important to keep mind that the business practices we study are in a domain, advertising, for which past work suggests principal-agent problems may be particularly large. The effects of advertising are hard to measure, and nominal metrics can distort the real impact of campaigns (Lewis and Rao, ming; Lewis et al., 2015). Marketing managers are often assigned to a certain class of media, *e.g.* television or digital, and may face negative personal consequences if disappointing effectiveness results are found in their domain. Viewed in this light, BNT represented *good news* for a firm advertising on own-brand keywords—it may be able to save millions of dollars just as eBay did. But it probably represented *bad news* for the people making these decisions "on the ground" because past expenditure could be revealed as wasteful, and budgets cuts, loss of positions and so forth may follow. Seen from this angle, new information is risky, which also helps explain why we do not observe increases in experimentation. Fundamental changes in measurement methodology may take longer to take hold and could be driven by education in programs such as MBAs.

It is certainly not the case that all business practices are subject to agency problems of this nature. Medicine is perhaps at the other extreme. Doctors actively contribute to the academic literature, and published advances generally benefit doctors, hospitals and patients alike. Indeed, there are "evidenced based medicine" journals that have the mission to ensure the new methods are quickly adopted (Haynes et al., 2002). In addition, doctors are also highly trained to process technical information about the efficacy of new treatment options, and evidence suggests that their choices among competing treatments are broadly consistent with rational information processing with knowledge spillovers (Chandra and Staiger, 2007; Fiedler, 2013). While our findings do not extend to such cases where the incentive structures are entirely different, other authors have argued that the agency problems we discuss for advertising are generally present within large firms (*e.g.* Scharfstein and Stein (2000)). If so, our results may point to a broader drag on economic efficiency and could be an important factor in the wide variation in worker productivity observed in narrowly defined industries.

References

Ashraf, N., Bandiera, O., and Jack, B. K. (2014). No margin, no mission? a field experiment on incentives for public service delivery. *Journal of Public Economics*, 120:1–17.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *The Quarterly Journal of Economics*, pages 917–962.

Bandiera, O., Barankay, I., and Rasul, I. (2007). Incentives for managers and inequality among workers: Evidence from a firm-level experiment*. *The Quarterly Journal of Economics*, 122(2):729–773.

Bandiera, O., Barankay, I., and Rasul, I. (2011). Field experiments with firms. *The Journal of Economic Perspectives*, 25(3):63–82.

Bartelsman, E. J. and Doms, M. (2000). Understanding productivity: lessons from longitudinal microdata. *Journal of Economic literature*, 38(3):569–594.

Blake, T., Nosko, C., and Tadelis, S. (Forthcoming). Consumer heterogeneity and paid search effectivness: A large scale field experiment. *Econometrica*, pages 1–26.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics*, 128(1):1–51.

Bloom, N. and Van Reenen, J. (2010). Why do management practices differ across firms and countries? *The Journal of Economic Perspectives*, 24(1):203–224.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in healthcare: evidence from the treatment of heart attacks. *The Journal of Political Economy*, 115:103.

Edelman, B. and Ostrovsky, M. (2007). Strategic bidder behavior in sponsored search auctions. *Decision support systems*, 43(1):192–198.

Edelman, B., Ostrovsky, M., and Schwarz, M. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *The American Economic Review*, 97(1):242–259.

Fiedler, M. A. (2013). Essays on provider behavior in health care markets. *Harvard Department of Economics*.

Foster, L., Haltiwanger, J., and Syverson, C. (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review*, 98(1):394–425.

Haynes, R. B., Devereaux, P. J., and Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *Evidence Based Medicine*, 7(2):36–38.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, pages 1346–1361.

Lewis, R. A. and Rao, J. M. (Forthcoming). The unfavorable economics of measuring the returns to advertising. *Quarterly Journal of Economics*.

Lewis, R. A., Rao, J. M., and Reiley, D. H. (2015). chapter Measuring the Effects of Advertising: The Digital Frontier. NBER Press.

Ostrovsky, M. and Schwarz, M. (2011). Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 59–60. ACM.

Scharfstein, D. S. and Stein, J. C. (2000). The dark side of internal capital markets: Divisional rent-seeking and inefficient investment. *The Journal of Finance*, 55(6):2537–2564.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *The Review of Economic Studies*, 71(2):513–534.

Shin, H. S. (1994). News management and the value of firms. *The RAND Journal of Economics*, pages 58–71.

Simonov, A., Nosko, C., and Rao, J. M. (2017). Competition and crowd-out for brand keywords in sponsored search. *Working Paper*.

Yang, S. and Ghose, A. (2010). Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, 29(4):602–623.

6 Appendix

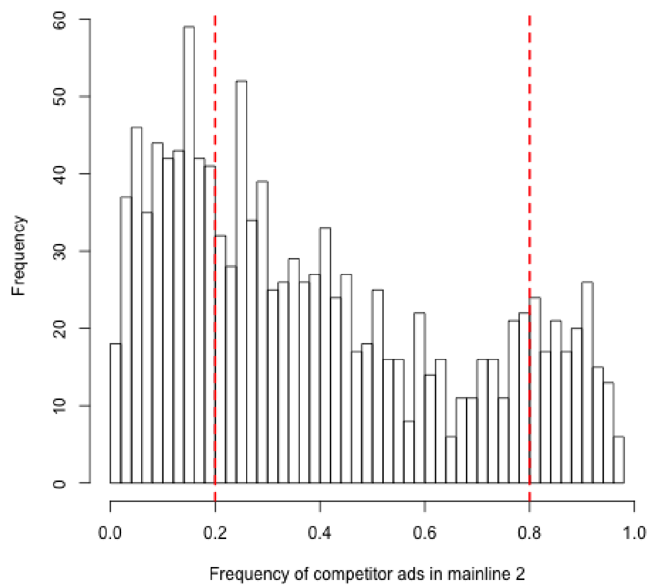
6.1 Press Coverage of Blake, Nosko and Tadelis (2015)

Major press includes:

Harvard Business Review, 3/11/2013. Did eBay Just Prove That Paid Search Ads (shown below)
Business Insider, 3/14/2013 EBay Slams Google Ads As A Waste Of Money
BBC, 3/13/2013. Google advertising value questioned by eBay.
The Economist, 7/13/2013. Simple tests can overstate the impact of search advertising.
The Atlantic, 4/13/2014. A dangerous question: Does internet advertising work at all?

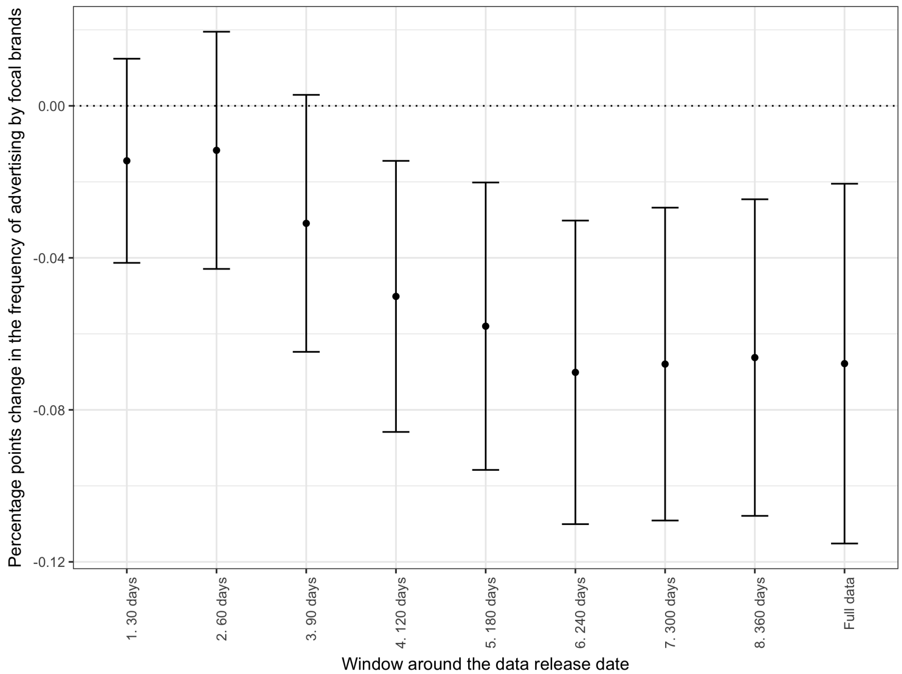
6.2 Frequency of competitors' ads in mainline 2

Figure 11: Histogram of Competitor's Ad Frequency in Mainline 2.



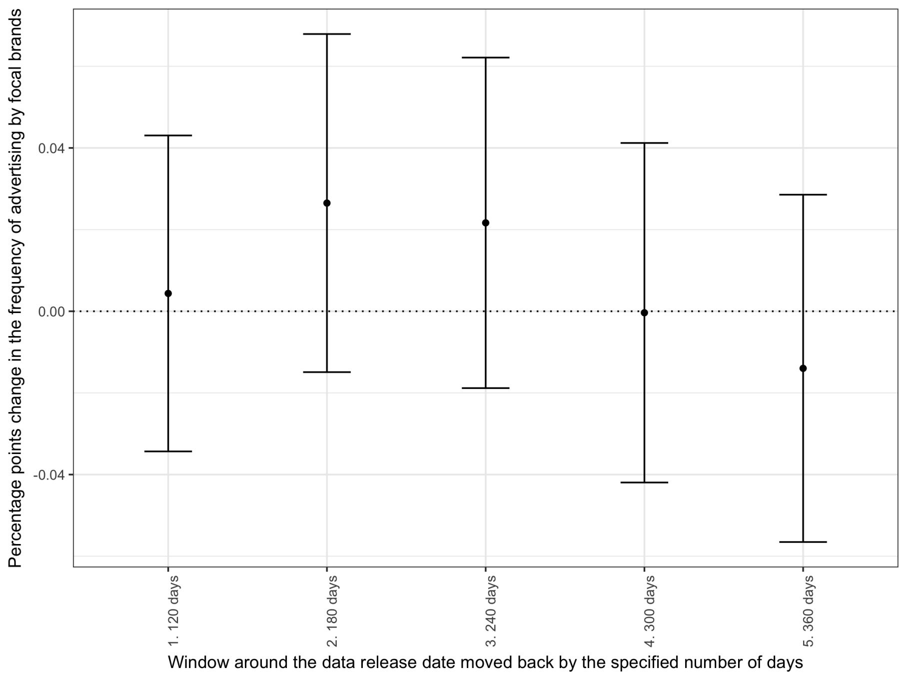
6.3 Advertising Level Reaction Timing and Placebo Tests

Figure 12: Estimates of the "treated" firms reaction to the BNT coverage, by time window around the information event.



The point estimates are the "Treatment * After" estimates from the specification (1) in Table 2 under different time windows.

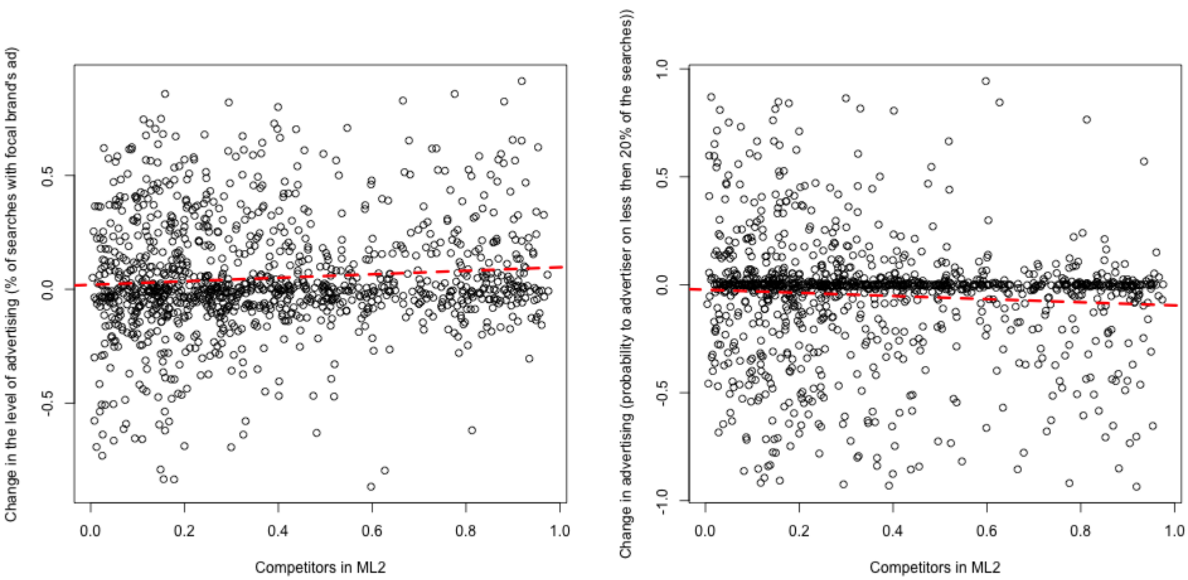
Figure 13: Placebo estimates of the "treated" firms reaction to the BNT coverage.



The point estimates are the "Treatment * After" estimates from the specification (1) in Table 2 under different time windows. The information event is moved back the same number of days as in the specified event window.

6.4 Advertising Level Reaction by Level of Competition

Figure 14: Own advertisement level for companies by the level of competition.

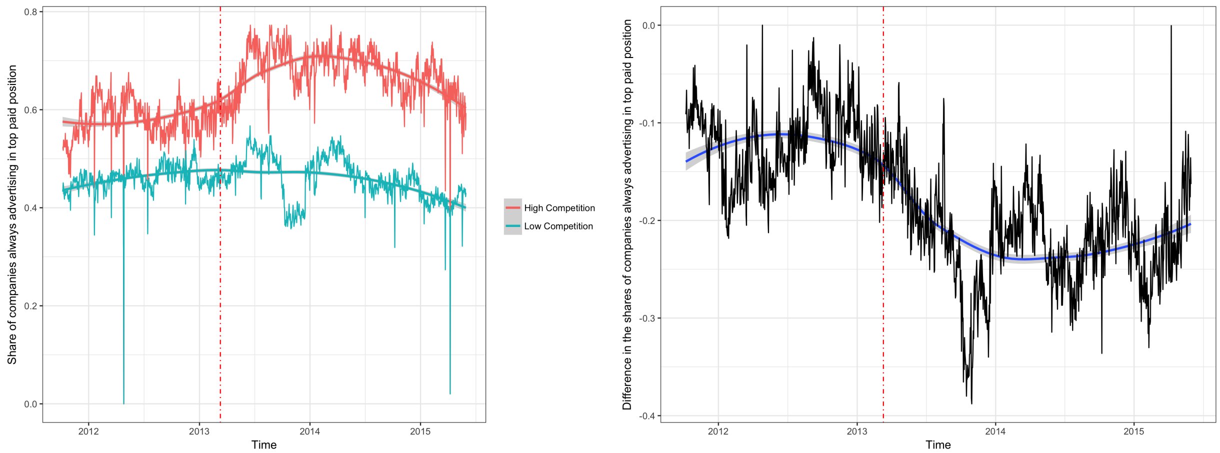


(a) Average Advertising Level (p-value = 0.004) (b) Probability to Advertise (p-value = 0.0121)

Based on 1148 firms in the sample. Each dot represents a firm. The dotted line represents a slope coefficient in the linear regression of the change in the level of advertising on the level of competition. Both slope coefficients are significant at 5% level: for (a) 0.077 (s.e. of 0.026), for (b) -0.073 (0.0289). Subfigure (a) corresponds to the results in Figure 5, Subfigure (b) – to the results in Figure 6.

6.5 Probability to Advertise More than 90% of the Time

Figure 15: Own advertisement level for “treatment” and “control” companies: fraction of companies advertising more than 90% of the time.

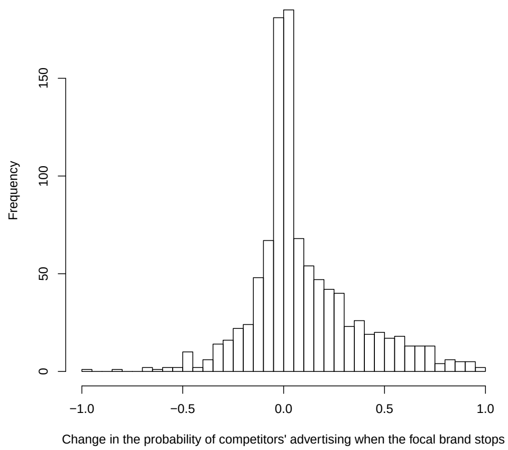


(a) Levels (b) Difference

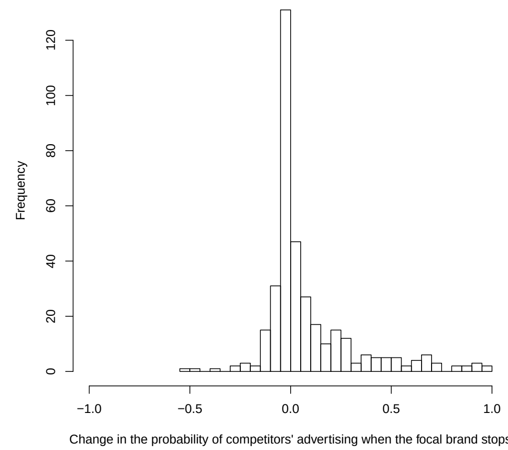
Fraction of companies with own brand ad in Mainline 1 > 90% of the time.

6.6 Competitors' Entry

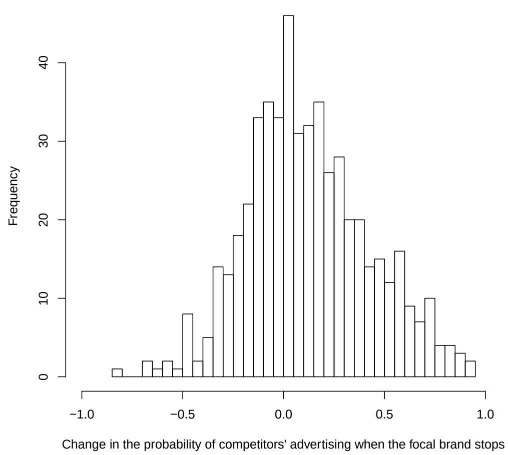
Figure 16: Competitors are more likely to advertise when the focal brand does not.



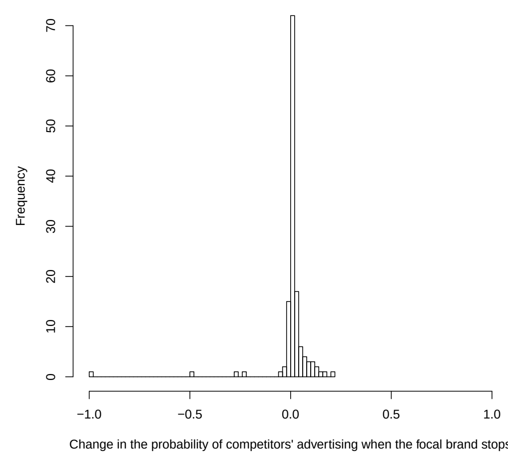
(a) All



(b) Low Competition



(c) Medium Competition

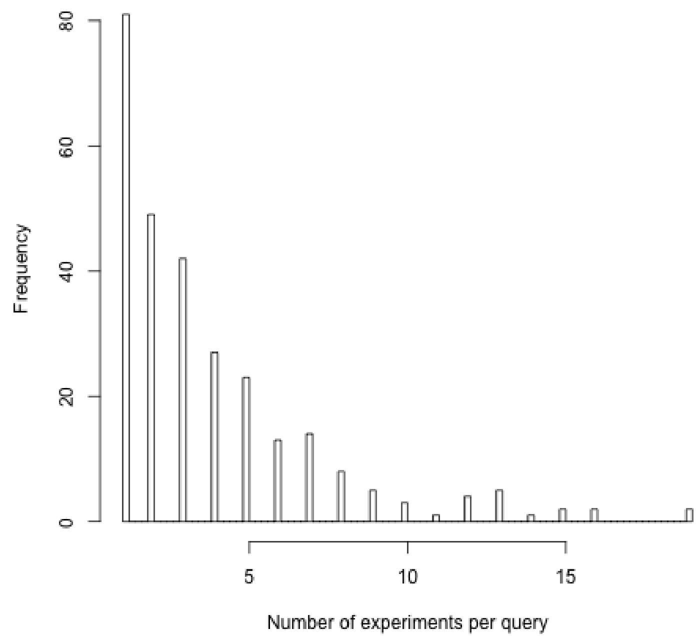


(d) High Competition

The histogram is across focal brands. Low, medium and high competition groups are defined based on the histogram 11. Competitive entry is defined as competitors advertising in paid position 1 or 2 at least 50% of the time. The probability of competitors' advertising is computed on a daily basis and averaged within a focal brand and days when the focal brand advertisers in paid position one more/less than 10% of the times.

6.7 Histogram of the Number of Experiments per Focal Nrand

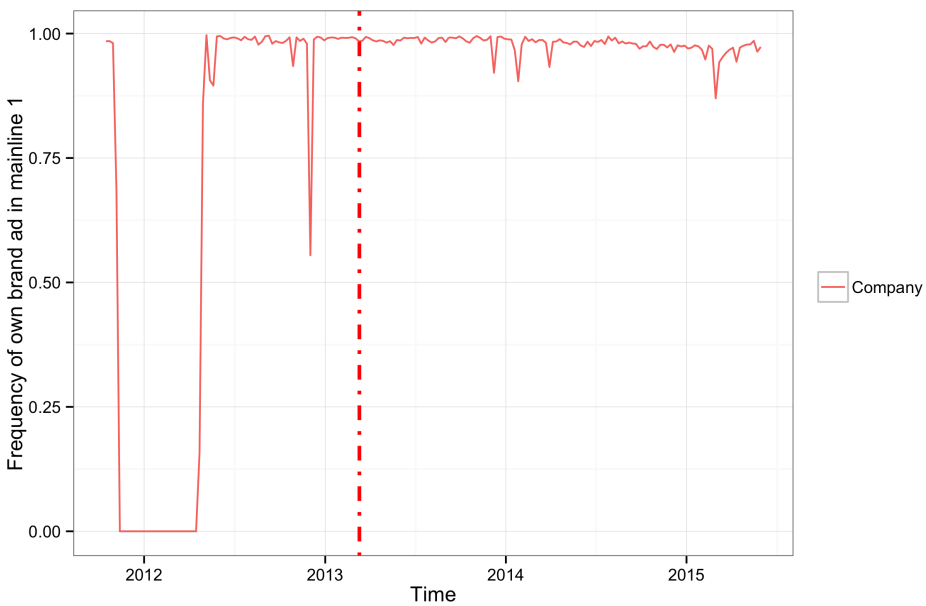
Figure 17: Histogram of the number of experiments per focal brand.



6.8 Example of an Experimenting Company

In Figure 18 we show an unnamed firm's advertising levels on own-brand keywords. This firm displayed a pattern of experimentation in the end of 2011 and beginning of 2012. Based on this experimentation, we find that advertising effect small and statisticaly insignificant. Nevertheless, this firm continued to advertise after the "experimentation" period.

Figure 18: Company X advertising frequency over time.



6.9 Estimated Effects versus Post-experimentation Behavior

Figure 19: Relationship between estimated effects for companies and their decision in the post-experimentation period. Done for companies with similar experimentation patterns as in the case of eBay.