Econometrica, Vol. 83, No. 1 (January, 2015), 155–174

CONSUMER HETEROGENEITY AND PAID SEARCH EFFECTIVENESS: A LARGE-SCALE FIELD EXPERIMENT

BY THOMAS BLAKE, CHRIS NOSKO, AND STEVEN TADELIS[1]

Internet advertising has been the fastest growing advertising channel in recent years, with paid search ads comprising the bulk of this revenue. We present results from a series of large-scale field experiments done at eBay that were designed to measure the causal effectiveness of paid search ads. Because search clicks and purchase intent are correlated, we show that returns from paid search are a fraction of non-experimental estimates. As an extreme case, we show that brand keyword ads have no measurable short-term benefits. For non-brand keywords, we find that new and infrequent users are positively influenced by ads but that more frequent users whose purchasing behavior is not influenced by ads account for most of the advertising expenses, resulting in average returns that are negative.

KEYWORDS: Advertising, field experiments, causal inference, electronic commerce, return on investment, information.

1. INTRODUCTION

ADVERTISING EXPENSES ACCOUNT for a sizable portion of costs for many companies across the globe. In recent years, the Internet advertising industry has grown disproportionately, with revenues in the United States alone totaling \$36.6 billion for 2012, up 15.2 percent from 2011. Of the different forms of Internet advertising, paid search advertising, also known in industry as "search engine marketing" (SEM), remains the largest advertising format by revenue, accounting for 46.3 percent of 2012 revenues, or \$16.9 billion, up 14.5 percent from \$14.8 billion in 2010. Google Inc., the leading SEM provider, registered \$46 billion in global revenues in 2012, of which \$43.7 billion, or 95 percent, were attributed to advertising.[2]

This paper reports the results from a series of controlled experiments conducted at eBay Inc., where large-scale SEM campaigns were randomly executed across the United States. The experiments show that the effectiveness of SEM is small for a well-known company like eBay and that the channel has been ineffective on average. We find, however, a significant positive effect of SEM on new user acquisition and on influencing purchases by infrequent and less recent users. This supports the *informative view* of advertising and implies that targeting *uninformed* users is a critical factor for successful advertising.

[1]We are grateful to many eBay employees and executives who made this work possible. We thank Susan Athey, Randall Lewis, Justin Rao, David Reiley, Florian Zettelmeyer, an editor, and three anonymous referees for comments on earlier drafts.

[2]See the Interactive Advertising Bureau (IAB) 2012 Full Year Results, April 2013, http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2012_rev.pdf and Google's webpage http://investor.google.com/financial/tables.html.

 DOI: 10.3982/ECTA12423

The effects of advertising on business performance have always been considered hard to measure. Traditional advertising channels such as TV, radio, print, and billboards have limited targeting capabilities, causing advertisers to waste valuable marketing dollars on "infra-marginal" consumers who are not affected by ads to get to those marginal consumers who are. The advent of Internet marketing channels has been lauded as the answer to this long-standing dilemma for two reasons.

First, the Internet lets advertisers target their ads to the activity that users are engaged in (Goldfarb (2014)). For instance, when a person reads content related to sports, like ESPN.com, advertisers can bid to have display ads appear on those pages. Similarly, if a user searches Google for information about flat-screen TVs, retailers and manufacturers of these goods can bid for paid search ads that better target the user's intent.

Second, advertisers can track data needed to measure the efficacy of ads because they will receive detailed data on visitors who were directed to their websites by the ad, how much was paid for the ad, and whether the visitor purchased anything from their website. This should allow the advertiser to compute the returns on investment.

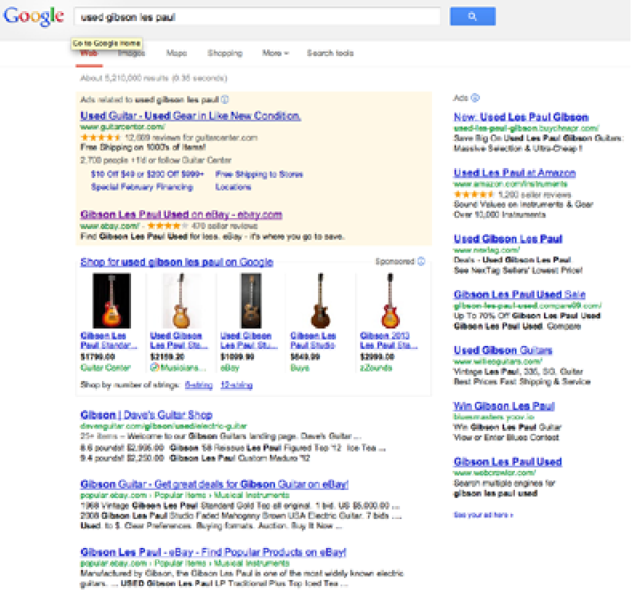
Despite these advantages, serious challenges persist in measuring causal relationships between Internet-advertising expenditures and sales. Traditionally, economists have focused on endogeneity stemming from firm decisions to increase advertising during times of high demand.[3] Our concern, instead, is that the amount spent on Internet marketing is a function not only of the advertiser's campaign, but is also determined by the *behavior* and *intent* of consumers because expenditures increase with clicks. In contrast, the amount spent on an ad in the *New York Times* print edition is independent of consumer behavior.

Our research highlights one potential drawback inherent in this form of targeting: In many cases, the consumers who choose to click on ads are loyal customers or are otherwise already informed about the company's product. Advertising may appear to attract these consumers, when in reality they would have found other channels to visit the company's website. We overcome this endogeneity challenge with our controlled experiments.

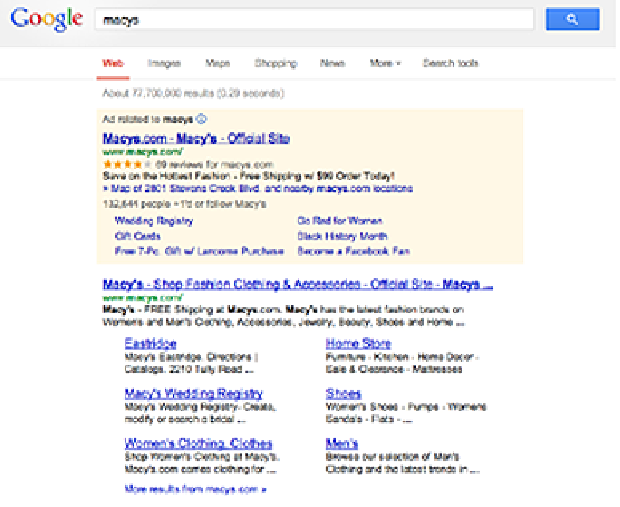
To visualize SEM, Figure 1(a) shows Google search results for the query "used gibson les paul," which fall into two categories: paid (or "sponsored") search ads (two on the top, five photo ads, and seven on the right), and unpaid (also called "natural" or "organic") search results (three at the bottom). Organic results are ranked by Google's "PageRank" algorithm, while the placement of paid search ads depends on bids made by advertisers when a particular query is typed by a user. Advertisers pay only when a user clicks on the ad, implying that ad expenses are only incurred for users who respond to the ad.

The example in Figure 1(a) describes what is referred to as a non-brand keyword search, despite the fact that a particular branded product (Gibson

[3]For example, advertising during the holidays, or when advertising budgets are set as a percentage of previous-quarter revenue. See Berndt (1991, Chapter 8), for a survey of this literature.



(a) Used Gibson Les Paul



(b) Macys

FIGURE 1.—Google ad examples. Panel (a) shows Google search results for a non-brand term "used gibson les paul" and panel (b) shows a brand term, Macy's.

Les Paul) is part of the query, because many retailers will offer this guitar for sale. This is in contrast to a branded keyword such as "macys." Figure 1(b) shows the results page from searching for "macys" on Google with only one paid ad that links to Macy's main webpage.

To illustrate a striking example of the endogeneity problem, we first tested the efficacy of brand keyword advertising, a practice used by many companies. For example, on February 16, 2013, Google searches for the keywords "AT&T," "Macy," "Safeway," "Ford," and "Amazon" resulted in paid ads at the top of the search results page directly above organic unpaid links to the companies' sites. Brand paid search links simply intercept consumers at the last point in their navigational process, resulting in an extreme version of the endogeneity concern because users would have found the advertisers' sites anyway.

Section 2 presents experiments showing that there is no measurable short-term value in brand keyword advertising. eBay halted SEM queries for brand keywords (all queries that included the term eBay, e.g., "ebay shoes") on both Yahoo! and Microsoft (MSN), while continuing to pay for these terms on Google, which we used as a control in our estimation. The results show that almost all of the forgone click traffic and attributed sales were captured by natural search.[4] That is, substitution between paid and unpaid traffic was nearly complete. We further confirm this result using several brand keyword experiments on Google's search platform.[5]

Section 3 presents our main analyses based on experiments for non-branded keyword advertising. eBay historically managed over 100 million keywords and keyword combinations using algorithms that are updated daily and automatically feed into Google's, Microsoft's, and Yahoo!'s search platforms.[6] Examples of such keyword strings are "memory," "cell phone," and "used gibson les paul." Unlike branded search, where a firm's website is usually in the top organic search slot, organic placement for non-branded terms varies widely. The question is whether, absent SEM ads, consumers will use other channels to navigate to eBay's website (e.g., by directly navigating to www.ebay.com).[7]

To address this question, we designed a controlled experiment using Google's geographic bid feature (see Vaver and Koehler (2011)) that can determine,

[4]Throughout, we refer to sales as the total dollar value of goods purchased by users on eBay. Revenue is close to a constant fraction of sales, so percentage changes in the two are almost equivalent.

[5]The distinction in Rutz and Bucklin (2011) between generic and branded search terms is similar to our distinction between brand and non-brand terms. They did not measure the efficacy of brand search ads.

[6]See "Inside eBay's business intelligence" by Jon Tullett, news analysis editor for ITWeb at http://www.itweb.co.za/index.php?option=com_content&view=article&id=60448:Inside-eBay-s-business-intelligence&catid=218.

[7]Yang and Ghose (2010) investigated whether organic and paid search links are substitutes or complements. They did not measure sales related to other channels or the total effect of SEM on sales.

with a reasonable degree of accuracy, the geographic area of the user conducting each query. We designated a random sample of 30 percent of eBay's U.S. traffic in which we stopped all bidding for all non-brand keywords for 60 days. The test was designed to estimate the effect of paid search on sales and allowed us to explore heterogeneous responses across a wide consumer base.

The experiment showed that SEM had a very small and statistically insignificant effect on sales. We then segmented users according to the frequency and recency at which they visit eBay. We found that SEM accounted for a statistically significant increase in new registered users and purchases made by users who bought infrequently. SEM did not have a significant effect on the purchasing behavior of consumers who bought more frequently. We calculate that the short-term returns on investment for SEM were negative because frequent eBay shoppers account for most of the sales attributed to paid search.

Our results support the *informative view* of advertising, which posits that advertising informs consumers of the characteristics, location, and prices of products and services that they may otherwise be ignorant about.[8] In particular, consumers who have completed several eBay transactions in the year before our experiment are likely to be familiar with eBay and are unaffected by SEM. In contrast, more new users sign up when they are exposed to SEM ads, and users who only purchased one or two items in the previous year increase their purchases when exposed to SEM.

These results echo findings in Ackerberg (2001) who showed, using a reduced form model, that consumers who were not experienced with a product were more responsive to ads than consumers who had experienced the product. To the best of our knowledge, we analyze the first large-scale field experiment that documents the causal response of consumers to changes in advertising differentiated by how informed these consumers were.[9]

We contribute to a growing literature that exploits rich Internet marketing data to explore how consumers respond to advertising.[10] Lewis and Reiley (2014b) examined a related endogeneity problem to the one we stress, which they call "activity bias," where people who are more active online will both

[8] Bagwell (2007) gave an excellent review of the economics literature on advertising. The *persuasive view* of advertising suggests that consumers who are exposed to persuasive advertising will develop a preference for the advertised product. Intuitively, SEM is an advertising medium that affects the information that people have, and is unlikely to play a persuasive role. It is possible that display ads, which appear on pages without direct consumer queries, may play more of a persuasive role. A few papers have explored the effects of display ads on offline and online sales. See Manchanda, Dubé, Goh, and Chintagunta (2006), Goldfarb and Tucker (2011), and Lewis and Reiley (2014b).

[9] Other recent papers have shown heterogeneous responses of consumers along demographic dimensions such as age, gender, and location. See Lewis and Reiley (2014a) and Johnson, Lewis, and Reiley (2014).

[10] See Yao and Mela (2011), Chan, Wu, and Xie (2011), Reiley, Li, and Lewis (2010), and Narayanan and Kalyanam (2011) for recent papers that study SEM using other methods.

see more display ads and click on more links.[11] To illustrate the severity of the endogeneity problem in our data, we calculated the Return on Investment (ROI) using typical ordinary least squares (OLS) methods, which result in a ROI of over 4,100% without time and geographic controls, and a ROI of over 1,600% with such controls. We then used our experimental methods to control for endogeneity and found a ROI of −63%, with a 95% confidence interval of [−124%, −3%], rejecting the hypothesis that the channel yields any short-run positive returns. This result further emphasizes the importance of using controlled experiments in measuring the effectiveness of advertising, a tradition going back to at least Lodish, Abraham, Livelsberger, Lubetkin, Richardson, and Stevens (1995).[12]

If, as we suspect, our results generalize to other well-known brands that are in most consumers' consideration sets, then our study suggests that much of what is spent on Internet advertising may be beyond the peak of its efficacy. We conclude by discussing the challenges that companies face in choosing optimal levels of advertising.

2. BRAND SEARCH EXPERIMENTS

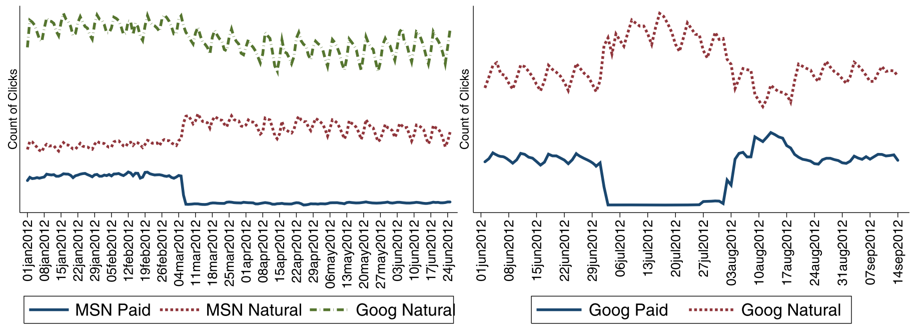
In March of 2012, eBay conducted a test to study the returns of brand keyword search advertising. Brand terms are any queries that *include* the term eBay such as "ebay shoes." Our hypothesis was that users who type "eBay" are using search as navigation with the intent to go to www.ebay.com. If so, brand ads "intercept" those searches because the natural search results will serve as an almost perfect substitute. To test this hypothesis, eBay halted advertising for its brand related terms on Yahoo! and MSN. The experiment revealed that almost all (99.5 percent) of the forgone click traffic from turning off brand keyword paid search was immediately captured by natural search traffic from the platform, in this case Yahoo! and MSN (Bing). That is, substitution between paid and unpaid traffic was nearly complete.[13]

Figure 2(a) plots the paid and natural clicks originating from the search platform. Paid clicks were driven to zero when advertising was suspended, while there was a noticeable uptake in natural clicks. Since users intend to find eBay, it is not surprising that shutting down the paid search path to their desired des-

[11]Edelman (2013) raised the concern that industry measurement methods, often referred to as "attribution models," may indeed overestimate the efficacy of such ads. Lewis and Rao (2013) exposed another problem with measurement showing that there are significant problems with the power of many experimental advertising campaigns, leading to wide confidence intervals.

[12]Other recent papers that use controlled experiments to investigate related issues include Sahni (2012), Lewis and Reiley (2014b), and Lambrecht and Tucker (2013).

[13]The 0.5 percent of all clicks lost represents about 1.5 percent of all paid clicks. In a recent paper, Yang and Ghose (2010) similarly switched off and back on paid search advertising for a random set of 90 keywords. We find much smaller differences in total traffic, most likely because we experimented with a brand term where the substitution effect is much larger.



(a) MSN Test

(b) Google Test

FIGURE 2.—Brand keyword click substitution. MSN and Google click-traffic counts to eBay on searches for 'ebay' terms are shown for two experiments where paid search was suspended (panel (a)) and suspended and resumed (panel (b)).

tination simply diverts traffic to the next easiest path, natural search, which is free to the advertiser.

To quantify this substitution, we first regressed the log of total daily clicks from MSN to eBay on an indicator for whether days were in the period with ads turned off. Click volume was 5.6 percent lower in the period after advertising was suspended. We then used data on eBay's clicks from Google as a control for seasonal factors because during the test period on MSN, eBay continued to purchase brand keyword advertising on Google. We performed a difference-in-differences analysis using Google as a control.[14] Once the seasonality is accounted for, only 0.529 percent of the click traffic is lost, so 99.5 percent is retained. Note that this is a lower bound of retention because some of the 0.5 percent of lost traffic may have switched to typing "ebay.com" into the browser.

These results inspired a follow-up test on Google's platform that was executed in July of 2012, which yielded similar results. Figure 2(b) shows both the substitution to natural traffic when search advertising was suspended and the substitution back to paid traffic when advertising resumed. In total, traffic referred by Google dropped by 3.2 percent. It is likely that a well-constructed control group would reduce this estimate, as was evident in the MSN test. During this test, there was no viable control group because there was no other contemporaneous paid search brand advertising campaign. In the Supplemental Material (Blake, Nosko, and Tadelis (2015)), we describe a test in Germany that preserved a control group, which confirms the results.

In summary, the evidence strongly supports the intuitive notion that for brand keywords, natural search is close to a perfect substitute for paid search,

[14]Detailed results are shown in Table A.I of the Supplemental Material (Blake, Nosko, and Tadelis (2015)).

making brand keyword SEM 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.

3. NON-BRAND TERMS CONTROLLED EXPERIMENT

When typing queries for non-brand terms, users may be searching for information on goods or wish to purchase them. If ads appear for users who do not know that these products are available at the advertiser's website, then there is potential to bring these users to the site, which in turn might generate sales that would not have occurred without the ads.

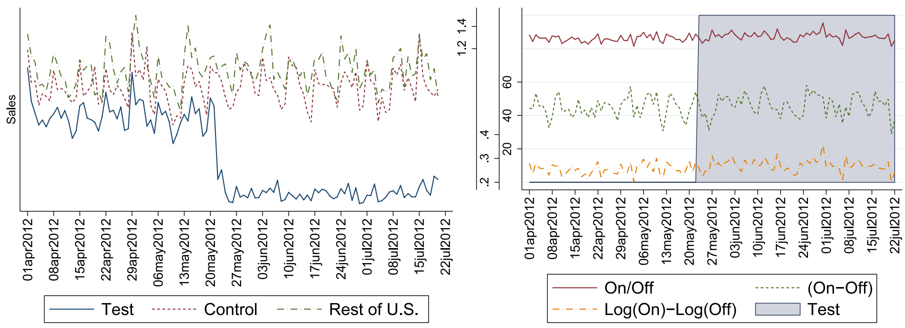
Because eBay bids on a universe of over 100 million keywords, it provides an ideal environment to test the effectiveness of paid search ads for non-brand keywords. The broad set of keywords place ads in front of a wide set of users who search for queries related to millions of products. Measuring the effects of the full keyword set more directly addresses the value of informative advertising because we can examine how consumers with different levels of familiarity with the site respond to advertising. In particular, we can use measures of past activity on eBay to segment users into groups that would be more or less familiar with eBay's offerings. Non-brand ads can attract users who are not directly searching for eBay, but the endogeneity problem persists because the ads may attract informed users who may have visited eBay even if the ad were not present.

3.1. *Experimental Design and Basic Results*

To measure the effect of advertising on non-brand queries, we implemented a large-scale field experiment that exposed a random group of users to ads, while a control group did not see ads.[15] We used Google's geographic bid feature that determines, with a reasonable degree of accuracy, the Nielsen Designated Market Area (DMA) of the user conducting each query. There are 210 DMAs in the United States, which typically correspond to large metropolitan areas. For example, San Francisco, Oakland, and San Jose, CA, comprise a large DMA, while Butte and Bozeman, MT, comprise a smaller DMA.

For the test, ads were suspended in roughly 30 percent of DMAs. This was done to reduce the scope of the test and minimize the potential cost and impact to the business (in the event that the ads created considerable profits). A purely random subsample of DMAs were chosen as candidates for the test. Next, candidate DMAs were divided into test and control DMAs using an algorithm that

[15] Whereas the previous section referred to a test of advertising for branded keywords and their variants, this test specifically excluded brand terms. That is, eBay continued to purchase brand ads nationally until roughly 6 weeks into the geographic test, when the brand ads were halted nationwide.



(a) Attributed Sales by Region

(b) Differences in Total Sales

FIGURE 3.—Non-brand keyword region test. Panel (a) plots total purchases by users who clicked on an ad prior to purchase, which drops when the test commences in the test areas. Panel (b) plots three different measures of the difference between test and control regions before and after the test. The y-axis is shown for the ratio, the log difference, and in differences in thousands of dollars per day, per DMA.

matched historical serial correlation in sales between the two regions. This was done to create a control group that mirrored the test group in seasonality. This procedure implies that the test group is not a purely random sample, but it is certainly an arbitrary sample that does not exhibit any historical (or, ex post) difference in sales trends. The test design therefore lends itself neatly to a standard difference-in-differences estimation of the effect of paid search on sales. For robustness, we explored alternative estimation strategies using the purely random assignment in the Supplemental Material.

Figure 3(a) plots total *attributed* sales for the three regions of the United States: the 68 test DMAs where advertising ceased, 65 matched control DMAs, and the remaining 77 control DMAs (yielding a total of 142 control DMAs). Attributed sales are total sales of all purchases by users within 24 hours of that user clicking on a Google paid search link.[16] Note that attributed sales did not completely zero out in the test DMAs during the experiment (they fell by over 72 percent). The remaining ad sales from test DMAs are an artifact of the error both in Google's ability to determine a user's location and our determination of the user's location. We used the user's shipping zip code registered with eBay to determine the user's DMA and whether or not the user was exposed to ads. If a user made a purchase while traveling to a region exposed to ads but still had the product shipped to her home, we would assign the associated sales to the off region.[17]

[16]The y-axis is suppressed to protect proprietary sales data. It is in units of dollars per DMA, per day.

[17]This classification error will attenuate the estimated effect towards zero. However, the Instrumental Variables estimates in Columns (3) and (4) of Table I measure an effect on the intensive margin of spending variation, which overcomes the classification error problem.

Figure 3(b) plots the simple difference, ratio, and log difference between daily average sales in the designated control regions where search remained on and the test regions where search was off. The regions where search remained on are larger (about 30 percent) than the regions switched off as a result of the selection algorithm that optimized for historical trends. There is no noticeable difference between the pre- and post-experimental period, demonstrating the muted overall effect of paid search. The Supplemental Material contains details of a difference-in-differences estimation that mimics Figure 3(b) using daily data from the full national set of DMAs in a regression of sales on indicators for whether or not paid search was turned on. Column (5) of Table I reports the results showing that the entire regime of paid search adds only 0.66 percent to sales with a 95 percent confidence interval of $[-0.42\%, 1.74\%]$.[18]

We now examine the magnitude of the endogeneity problem. Absent endogeneity problems, we could estimate the effect of ad spending on sales with a simple regression:

$$\ln(Sales_{it}) = \alpha_1 \times \ln(Spend_{it}) + \varepsilon_{it}, \tag{1}$$

where i indexes the DMA and t indexes the day. Columns (1) and (2) of Table I show the estimates of such a regression during the period prior to our test. As is evident, the simple OLS in Column (1) yields unrealistic returns suggesting that every 10 percent increase in spending raises revenues by 9 percent. The inclusion of DMA and day controls in Column (2) lowers this estimate to 1.3 percent, which is still very high. The amount spent on ads, however, depends on the search behavior of users, which is correlated with their intent to purchase. Our experiment overcomes this endogeneity problem.

Columns (3) and (4) of Table I instrument for spending with dummies for the experiment regions, experiment period, and interaction. We used a two stage least squares estimation with the following first stage:

$$\ln(Spend_{it}) = \tilde{\alpha}_1 \times AdsOn_{it} + \tilde{\alpha}_2 \times Post_t + \tilde{\alpha}_3 \times Group_i + \varepsilon_{it}, \tag{2}$$

where $Post_t$ is an indicator for whether the test was running, $Group_i$ is an indicator equal to 1 if region i kept search spending on, and $AdsOn_{it}$ is the interaction of the two indicators. The instruments isolate the exogenous experimental variation in spending to estimate the causal effect of spending on changes in revenue. True returns are almost two orders of magnitude smaller and are no longer statistically different from zero.

3.2. *Consumer Response Heterogeneity*

The scale of our experiment allows us to separate outcomes by observable user characteristics. Econometrically, this can be accomplished by interacting

[18]Alternate specifications are presented in Table A.II of the Supplemental Material.

TABLE I

RETURN ON INVESTMENT[a]

	OLS		IV		DnD	
	(1)	(2)	(3)	(4)	(5)	
Estimated Coefficient	0.88500	0.12600	0.00401	0.00188	0.00659	A
(Std Err)	(0.0143)	(0.0404)	(0.0410)	(0.0016)	(0.0056)	
DMA Fixed Effects		Yes		Yes	Yes	
Date Fixed Effects		Yes		Yes	Yes	
N	10,500	10,500	23,730	23,730	23,730	
$\Delta \ln(Spend)$ Adjustment	3.51	3.51	3.51	3.51	1	B
$\Delta \ln(Rev)$ (β)	3.10635	0.44226	0.01408	0.00660	0.00659	$\mathrm{C} = \mathrm{A} * \mathrm{B}$
Spend (Millions of $)	$51.00	$51.00	$51.00	$51.00	$51.00	D
Gross Revenue (R′)	2,880.64	2,880.64	2,880.64	2,880.64	2,880.64	E
ROI	4,173%	1,632%	−22%	−63%	−63%	$\mathrm{F} = \mathrm{A}/(1+\mathrm{A}) * (\mathrm{E}/\mathrm{D}) - 1$
ROI Lower Bound	4,139%	697%	−2,168%	−124%	−124%	
ROI Upper Bound	4,205%	2,265%	1,191%	−3%	−3%	

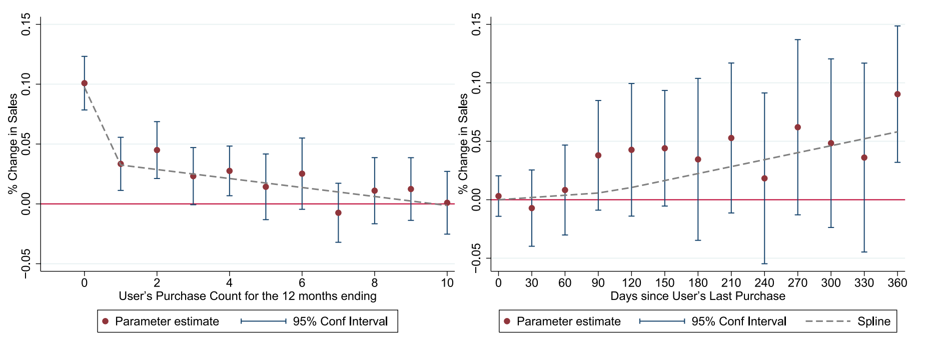
[a]The upper panel presents regression estimates of SEM's effect on sales. Columns (1) and (2) naively regress sales on spending in the pre-experiment period. Columns (3) and (4) show estimates of spending's effect on revenue using the difference-in-differences indicators as excluded instruments. Column (5) shows the reduced form difference-in-differences interaction coefficient. The lower panel translates these estimates into a return on investment (ROI) as discussed in Section 4 and shows its 95% confidence interval.

the treatment dummy with dummies for each subgroup, which produces a set of coefficients representing the total average effect from the advertising regime on that subgroup. We examined user characteristics that are common in the literature: the recency and frequency of a user's prior purchases. First, we interacted the treatment dummy with indicators for the number of purchases by that user in the year before April 2012. We estimated the following specification:

$$\ln(Sales_{imt}) = \beta_m \times AdsOn_{imt} \times \theta_m + \delta_t + \gamma_i + \theta_m + \varepsilon_{it}, \tag{3}$$

where $m \in \{0, 1, \dots, 10\}$ indexes user segments. Users with no purchases in the prior year are indexed by $m = 0$, those who purchased once in the prior year by $m = 1$, and so on, while $Sales_{imt}$ is the total sales by all users in segment m in period t and DMA i. This produces 11 estimates, one for each user segment.[19] Figure 4(a) plots the point estimates of the treatment interactions. The largest effect on sales was for users who had not purchased before on eBay. Interestingly, the treatment effect diminishes quickly with purchase frequency and estimates are near zero for users who buy more regularly.[20]

Second, Figure 4(b) plots the interactions by time since last purchase. Estimates become noisier as we look at longer periods of inactivity because there



(a) User Frequency (b) User Recency

FIGURE 4.—Paid search effect by user segment. Panel (a) shows difference-in-differences estimates and 95% confidence bands of paid search effects on sales for different user segments as defined by how many purchases were made in the previous 12 months. Panel (b) shows similar estimates where users were segmented by the time since last purchase.

[19]This is similar to running 11 separate regressions, which produces qualitatively similar results.

[20]Only the zero purchases effect is statistically distinguishable from other segments, even when pooled into larger buckets. The slope of the relationship between effect and frequency is statistically negative, however, even excluding the zero purchase users. We illustrate this with a simple spline beginning at one purchase, shown as the dashed line. The slope of the right segment of this line, derived by replacing θ_m in Equation (3) with a continuous purchase count, is estimated to be -0.0038 with a standard error of 0.0014.

are fewer buyers that return after longer absences. The estimates are tightly estimated zeros for zero days and consistently centered on zero for 30 and 60 day absences, suggesting that advertising has little effect on active and moderately active customers. However, the effect then steadily rises with absence and becomes large and statistically significant for customers who have not purchased in over a year.[21] We estimated a spline with a break at the arbitrarily chosen 90 day mark and the estimate of the treatment effect is 0.02 percentage points larger per month of absence.[22]

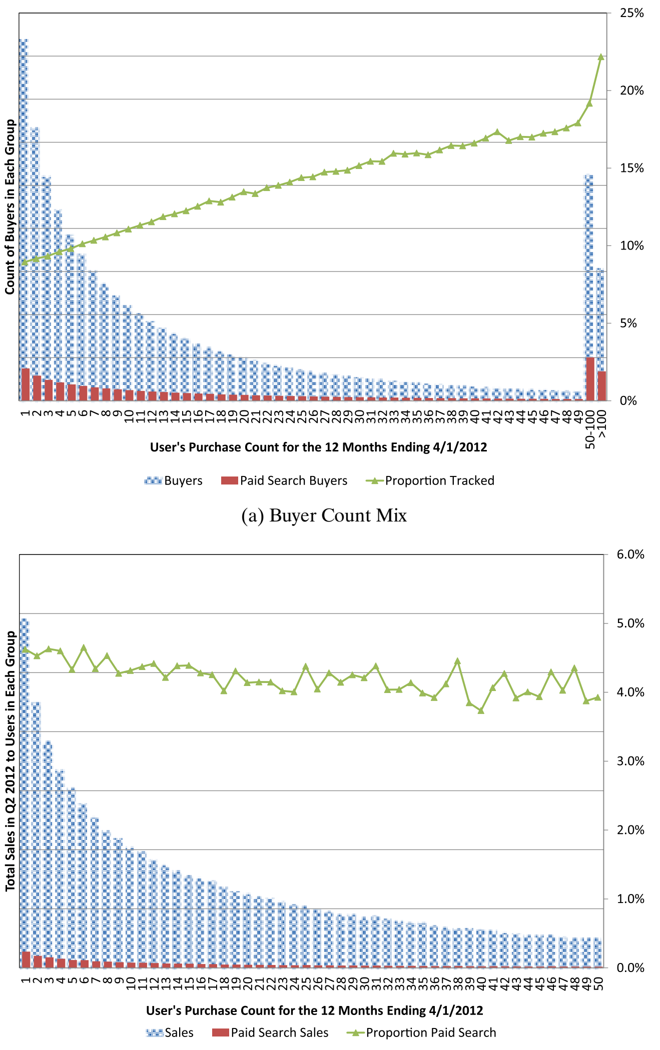
Figure 4 implies that search advertising works only on a firm's least active customers. These are traditionally considered a firm's "worst" customers, and advertising is often aimed at high value repeat consumers (Fader, Hardie, and Lee (2005)). Our evidence supports the informative view where ads affect consumption only when they update a consumer's information set. Bluntly, search advertising only works if the consumer has no idea that the company has the desired product. Large firms like eBay with powerful brands will see little benefit from paid search advertising because most consumers already know that they exist, as well as what they have to offer. The modest returns on infrequent users likely come from informing them that eBay has offerings they did not think were available.

While the least active customers are the best targets for search advertising, we find that most paid search traffic and attributed sales are from high volume, frequent purchasers. Figure 5(a) plots the count of buyers by how many purchases they made in a year. The counts are shown separately for all buyers and for those that buy, at any point in the year prior to the experiment, *after* clicking on a paid search ad. The ratio of the two rises with purchase frequency because frequent purchasers are more likely to use paid search at some point. Figure 5(b) shows the same plot for shares of transaction counts. Even users who buy more than 50 times in a year still use paid search clicks for 4 percent of their purchases. The large share of heavy users suggests that most of paid search spending is wasted because the majority of spending on Google is related to clicks by those users that would purchase anyway. This explains the large negative ROI computed in Section 4.

We have searched for other indicators of consumers' propensity to respond in localized demographic data. We interacted zip code demographic data with our treatment indicator and found no response that is statistically different across several demographic measures: income, population size, unemployment rates, household size, and eBay user penetration. Finally, we looked for differ-

[21]Gönül and Shi (1998) studied a direct mail campaign and found that recent individuals are not influenced by mailing because they are likely to buy anyway.

[22]This estimate is derived in similar fashion to the spline in panel (a): with interactions of the treatment dummy and the number of days since purchase. This is statistically distinguishable from zero, with a standard error of 0.00004577 so that pooling across user segments provides better evidence of the trend than the noisier separate coefficients.



(a) Buyer Count Mix

(b) Transaction Count Mix

FIGURE 5.—Paid search attribution by user segment. Panel (a) shows the histogram of all buyers by how many purchases were made in the previous 12 months compared to the subset of buyers with paid search clicks preceding purchases. The proportion tracked is the ratio of the two counts in each bucket. Panel (b) shows the same for the distribution of transaction counts.

ential effects across product types and found no systematic difference across multiple layers of categorization. See the Supplemental Material for more details.

3.3. *Where Did the Non-Brand Traffic Go?*

The brand query tests demonstrated that causal (incremental) returns were small because users easily substituted paid search clicks for natural search clicks. Metaphorically, we closed one door and users simply switched to the next easiest door. This substitution was expected because users were using brand queries as simple navigational tools. Unbranded queries are not simply navigational because users are trying to find *any* destination that has the desired product. Only experimental variation can quantify the number of users who are actually directed by the presence of search advertising.

Experimentation can also quantify the substitution between SEM and other channels. For example, in Figure 1(a) we showed Google's search results page for the query "used gibson les paul." Notice that the second ad from the top, as well as the center image of the guitar below it, are both paid ads that link to eBay, while the two bottom results of the natural search part of the page also link to eBay. Hence, some substitution from paid to natural search may occur for non-brand keywords as well. Also, users who intend to visit eBay and do not see ads may choose to directly navigate to www.ebay.com.

Clicks to eBay decline measurably in the absence of non-brand ads.[23] Advertising clicks dropped 41 percent, and total clicks fell 2 percent as a result of the non-brand experiment.[24] The total loss in clicks is roughly 58 percent of the number of lost paid search clicks, suggesting that 42 percent of paid search clicks are newly acquired. Advertising does increase clicks above and beyond what is taken from natural search. This mirrors studies from Google that find that the majority of lost paid search clicks would not have been recouped by natural search (Chan, Yuan, Koehler, and Kumar (2011)).

But clicks are just part of what generates sales. To make meaningful statements about Internet traffic, we need to make an important distinction in the nature of visits. eBay servers are able to distinguish between referring *unique clicks* (unique clicks from other sites that lead to an eBay page) and total *visits* (clusters of unique clicks by the same user). In the course of a single shopping session, users will have many unique clicks referring from other websites because their search takes them on and off eBay pages. Put simply, users will travel to eBay from Google multiple times in one sitting.

We defined a paid search visit as a session that begins with a paid search click and compare the substitution to comparably defined natural search visits. We measured the potential traffic substitution by regressing the log of eBay visit counts either from organic search or from direct navigation on the log of eBay visit counts from paid search, using the experiment as an instrument. We find that a 1 percent drop in paid search visits leads to a 0.5 percent increase in

[23]Recall that 99.5 percent of clicks were retained in the absence of brand paid ads.

[24]Natural clicks are a much larger denominator and therefore the total percentage drop is smaller.

natural search visits and to a 0.23 percent increase in direct navigation visits. These substitution results suggest that most, if not all, of the 'lost' traffic finds its way back through natural search and direct navigation. This helps explain why we found that clicks are lost but revenue is not.

4. DERIVING RETURNS ON INVESTMENT

To demonstrate the economic significance of our results, we computed the implied short-term return on investment (ROI) associated with spending on paid search. Imagine that the amount spent on paid search was S_0 associated with revenues equal to R_0. Let $\Delta R = R_1 - R_0$ be the difference in revenues as a consequence of an increase in spending, $\Delta S = S_1 - S_0$, and by definition, $ROI \equiv \frac{\Delta R}{\Delta S} - 1$.

Let $\beta_1 = \Delta \ln(R)$ be our estimated coefficient on paid search effectiveness, which is the effect of an increase in spend on log-revenues. (See the Supplemental Material for details on the estimation of β_1.) Using the definition of ROI and setting $S_0 = 0$ (no spending on paid search), some algebraic manipulation (detailed in the Supplemental Material) yields

$$ROI \approx \frac{\beta_1}{(1+\beta_1)} \frac{R_1}{S_1} - 1. \tag{4}$$

For the OLS and IV estimates where spending is the independent variable, we translate the coefficient $\alpha_1 = \frac{\Delta \ln(Sales)}{\Delta \ln(Spend)}$ from Equation (1) to a measure comparable to β_1 by multiplying by the coefficient $\tilde{\alpha}_1 = \Delta \ln(Spend)$ estimated from Equation (2), the first stage in the IV. This converts the IV estimates to reduced form estimates and approximates estimates derived from direct estimation of the difference-in-differences procedure. Both the derived and directly estimated β_1's can be used to compute a ROI with Equation (4).

In order to calculate the ROI from paid search, we need to use actual revenues and costs from the DMAs used for the experiment, but these are proprietary information that we cannot reveal. Instead, we used revenues and costs from public sources regarding eBay's operations. Revenue in the United States was derived from eBay's financial disclosures of Marketplaces' net revenue prorated to U.S. levels using the ratio of sales in the United States to global levels, which resulted in U.S. gross revenues of $2,880.64 million.[25] We next obtained paid search spending data from the release of information about the expenditures of several top advertisers on Google. We calculated eBay's yearly paid search spending for the United States to be $51 million.[26]

[25]Total revenues for 2012 were $7,398 and the share of eBay's activity in the U.S. was $26,424/$67,763 (in millions). See http://files.shareholder.com/downloads/ebay/2352190750x0x628825/e8f7de32-e10a-4442-addb-3fad813d0e58/EBAY_News_2013_1_16_Earnings.pdf.

[26]Data from Google report a monthly spend of $4.25 million, which we impute to be $51 million. See http://mashable.com/2010/09/06/brand-spending-google/.

Table I presents the ROI estimates. As is evident, simple OLS estimation of α_1 yields unrealistic returns of over 4,100 percent in Column (1) and even accounting for daily and geographic effects implies returns that are greater than 1,600 percent, as shown in Column (2). The IV estimation reduces the ROI estimate significantly below zero, and our best estimate of average ROI using the experimental variation is negative 63 percent as shown in Columns (4) and (5). This ROI is statistically different from zero at the 95 percent confidence level, emphasizing the economic significance of the endogeneity problem.

5. DISCUSSION

The efficacy of SEM seems limited at best for a well-known brand like eBay because expenditures are concentrated on consumers who would shop on eBay regardless of whether ads are shown. Of the $31.7 billion that was spent in the United States in 2011 on Internet advertising, estimates project that the top 10 spenders in this channel account for about $2.36 billion.[27] These companies generally use the same methods and rely on the same external support to design their ad campaigns, suggesting many reasons to believe that the results we presented above would generalize to these large and well-known corporations. This may not be true for small and new entities that have no brand recognition.[28]

This begs the question: why do well-known companies spend large amounts of money on what seems to be a rather ineffective marketing channel? One argument is that there are long-term benefits that we are unable to capture in our analysis. This does not seem to apply for brand keyword advertising because it is obvious that the user searched for the brand name and hence is aware of it. Arguments have been made that brand keyword advertising acts as a defense against a competitor bidding for a company's brand name. This implies that brand keyword advertising allows competing companies to play a version of the Prisoner's Dilemma. A company and its competitor would both be better off not buying any brand keywords, but each cannot resist the temptation to pinch away some of their competitor's traffic, and in the process, the ad platforms benefit from this rent-seeking game. It should be noted, however, that

[27]These include, in order of dollars spent, IAC/Interactive Group; Experian Group; GM; AT&T; Progressive; Verizon; Comcast; Capital One; Amazon; and eBay. See the press release by Kantar Media on 3/12/2012, http://kantarmediana.com/sites/default/files/kantareditor/Kantar_Media_2011_Full_Year_US_Ad_Spend.pdf.

[28]If you were to start a new online presence selling a high quality and low-priced widget, someone querying the word "widget" would still most likely not see your site. This is a consequence of the PageRank algorithm that relies on established links to webpages. Only after many websites link to your site, related to the word widget, will you stand a chance of rising to the top of the organic search results.

since eBay stopped bidding on its brand keywords, such behavior by potential competitors was not observed.[29]

Our experience suggests that one reason companies spend vast amounts on SEM stems from the challenges they face in generating causal measures of the returns to advertising. As the results in Table I demonstrate, typical regressions of sales on advertising spend result in astronomical ROI estimates that vastly overestimate the true ROI, which can only be estimated using controlled experiments. This is in line with results obtained by Lewis, Rao, and Reiley (2011) regarding the effectiveness of display ads.

In the absence of causal measures, the industry relies on 'attribution' measures which correlate clicks and purchases. By this measure, eBay performed very well, as shown in Figure 3(a) and Table I. eBay's ads were very effective at earning clicks and associated purchases. Our findings suggest, however, that even incremental clicks do not translate into incremental sales. This is an important way in which our methodology differs from the one used in studies released by Google. Chan et al. (2011) reported that experimental studies performed at Google proved that about 89% of paid search clicks were deemed to be incremental, that is, would not have happened if companies would not pay for search. As Section 3.3 shows, our results confirm that a majority of eBay's paid search clicks are not recovered when eBay stops paying for them. Nonetheless, the majority of these clicks did not result in incremental sales, which in turn is the reason that paid search was ineffective as clicks alone are not a source of revenues. It is interesting to note that the incentives faced by advertising firms, publishers, analytics consulting firms, and even marketing executives within companies, are all aligned with increasing advertising budgets.

REFERENCES

ACKERBERG, D. A. (2001): "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32 (2), 316–333. [159]

BAGWELL, K. (2007): "The Economic Analysis of Advertising," in *Handbook of Industrial Organization*, Vol. 3. Amsterdam: Elsevier, 1701–1844. [159]

BERNDT, E. R. (1991): *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: Addison-Wesley. [156]

BLAKE, T., C. NOSKO, AND S. TADELIS (2015): "Supplement to 'Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA12423. [161]

CHAN, D. X., Y. YUAN, J. KOEHLER, AND D. KUMAR (2011): "Incremental Clicks: The Impact of Search Advertising," *Journal of Advertising Research*, 51 (4), 643. [169,172]

CHAN, T. Y., C. WU, AND Y. XIE (2011): "Measuring the Lifetime Value of Customers Acquired From Google Search Advertising," *Marketing Science*, 30 (5), 837–850. [159]

[29]Some retailers have been bidding, both before and after eBay's response, to "broad" brand phrases such as "ebay shoes." It is also interesting to note that several advertisers pushed unsuccessfully for litigation to prevent their competitors from bidding on their trademark keywords, suggesting that some companies understand the Prisoner's Dilemma nature of this activity.

EDELMAN, B. (2013): "The Design of Online Advertising Markets," in *The Handbook of Market Design*. Oxford: Oxford University Press, 363. [160]

FADER, P. S., B. G. S. HARDIE, AND K. L. LEE (2005): "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, 42 (4), 415–430. [167]

GOLDFARB, A. (2014): "What Is Different About Online Advertising?" *Review of Industrial Organization*, 44 (2), 115–129. [156]

GOLDFARB, A., AND C. TUCKER (2011): "Online Display Advertising: Targeting and Obtrusiveness," *Marketing Science*, 30 (3), 389–404. [159]

GÖNÜL, F., AND M. Z. SHI (1998): "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science*, 44 (9), 1249–1262. [167]

JOHNSON, G. A., R. A. LEWIS, AND D. H. REILEY (2014): "Location, Location, Location: Repetition and Proximity Increase Advertising Effectiveness," Working Paper. [159]

LAMBRECHT, A., AND C. TUCKER (2013): "When Does Retargeting Work? Information Specificity in Online Advertising," *Journal of Marketing Research*, 50 (5), 561–576. [160]

LEWIS, R. A., AND J. M. RAO (2013): "On the Near Impossibility of Measuring the Returns to Advertising," Working Paper. [160]

LEWIS, R. A., AND D. H. REILEY (2014a): "Advertising Effectively Influences Older Users: How Field Experiments Can Improve Measurement and Targeting," *Review of Industrial Organization*, 44 (2), 147–159. [159]

——— (2014b): "Online Ads and Offline Sales: Measuring the Effects of Online Advertising via a Controlled Experiment on Yahoo!," *Quantitative Marketing and Economics*, 12 (3), 235–266. [159,160]

LEWIS, R. A., J. M. RAO, AND D. H. REILEY (2011): "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*. New York: ACM, 157–166. [172]

LODISH, L. M., M. M. ABRAHAM, J. LIVELSBERGER, B. LUBETKIN, B. RICHARDSON, AND M. E. STEVENS (1995): "A Summary of Fifty-Five In-Market Experimental Estimates of the Long-Term Effect of TV Advertising," *Marketing Science*, 14 (3_supplement), G133–G140. [160]

MANCHANDA, P., J.-P. DUBÉ, K. Y. GOH, AND P. K. CHINTAGUNTA (2006): "The Effect of Banner Advertising on Internet Purchasing," *Journal of Marketing Research*, 43 (1), 98–108. [159]

NARAYANAN, S. AND K. KALYANAM (2011): "Measuring Position Effects in Search Advertising: A Regression Discontinuity Approach," Working Paper. [159]

REILEY, D. H., S.-M. LI, AND R. A. LEWIS (2010): "Northern Exposure: A Field Experiment Measuring Externalities Between Search Advertisements," in *Proceedings of the 11th ACM Conference on Electronic Commerce*. New York: ACM, 297–304. [159]

RUTZ, O. J., AND R. E. BUCKLIN (2011): "From Generic to Branded: A Model of Spillover in Paid Search Advertising," *Journal of Marketing Research*, 48 (1), 87–102. [158]

SAHNI, N. (2012): "Effect of Temporal Spacing Between Advertising Exposures: Evidence From an Online Field Experiment," Working Paper, Graduate School of Business, Stanford University. [160]

VAVER, J., AND J. KOEHLER (2011): "Measuring Ad Effectiveness Using Geo Experiments," Technical Report, Google Inc. [158]

YANG, S., AND A. GHOSE (2010): "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence?" *Marketing Science*, 29 (4), 602–623. [158,160]

YAO, S., AND C. F. MELA (2011): "A Dynamic Model of Sponsored Search Advertising," *Marketing Science*, 30 (3), 447–468. [159]

eBay Research Labs, 2065 Hamilton Ave., San Jose, CA 95125, U.S.A.; thblake@ebay.com, *University of Chicago, 5807 South Woodlawn Ave., Chicago, IL 60637, U.S.A. and eBay Research Labs; cnosko@chicagobooth.edu,*

and

UC Berkeley, 2220 Piedmont Ave., Berkeley, CA 94720-1900, U.S.A., NBER, and eBay Research Labs; stadelis@haas.berkeley.edu.

Manuscript received April, 2014; final revision received August, 2014.