NBER WORKING PAPER SERIES

WILL AD BLOCKING BREAK THE INTERNET?

Ben Shiller
Joel Waldfogel
Johnny Ryan

Working Paper 23058
http://www.nber.org/papers/w23058

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2017

Johnny Ryan is an employee of PageFair, a firm that measures the extent of ad blocking and offers ad-recovery technologies that enable publishers to display ads in a manner that adblockers cannot block. Neither Shiller nor Waldfogel has received any compensation for the research in the paper. PageFair was allowed to review the paper for accuracy but did not have control over the paper's findings nor conclusions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Will Ad Blocking Break the Internet?
Ben Shiller, Joel Waldfogel, and Johnny Ryan
NBER Working Paper No. 23058
January 2017
JEL No. L81,L82

ABSTRACT

Ad blockers allow Internet users to obtain information without generating ad revenue for site owners; and by 2016 they were used by roughly a quarter of site visitors. Given the ad-supported nature of much of the web, ad blocking poses a threat to site revenue and, if revenue losses undermine investment, a possible threat to consumers' access to appealing content. Using unique, proprietary, and site-specific data on the share of site visitors using ad blockers at a few thousand sites, along with Alexa traffic data, we explore the impact of ad blocker usage on site quality, as inferred from traffic ranks, 2013-2016. We find that each additional percentage point of site visitors using ad blockers raises (worsens) its traffic rank by about 0.6 percent over a 35 month period, with stronger effects at initially worse-ranked sites. We provide additional evidence of causality by showing that the relationship between traffic trends and eventual ad blocking does not predate ad blocking. Plausible instruments for ad blocking also deliver consistent results. Effects of ad blocking on revenue are compounded by the fact that ad blocking reduces visits, while also generating less revenue from remaining visitors employing ad blockers. We conclude that ad blocking poses a substantial threat to the ad-supported web.

Ben Shiller
Brandeis University
Department of Economics
Waltham, MA 02453
shiller@brandeis.edu

Johnny Ryan
PageFair
johnny@pagefair.com

Joel Waldfogel
Frederick R. Kappel Chair in Applied Economics
3-177 Carlson School of Management
University of Minnesota
321 19th Avenue South
Minneapolis, MN 55455
and NBER
jwaldfog@umn.edu

Copyright protection is an important theoretical determinant of incentives for creation of new products, and recent research documents the relationship between legal intellectual property regime and the quality and quantity of new output.[1] Beyond legal rules alone, intellectual property is in general protected by a combination of law and technology that may together allow sellers to appropriate enough of consumers' willingness to pay to finance the creation of new products. The development of various digital technologies over the past few decades has undermined sellers' ability to appropriate consumers' willingness to pay, raising the possibility of a diminished supply of new information and entertainment products. For example, new digital technologies have challenged traditional revenue models through direct piracy (e.g. Napster), diversion of attention to aggregators rather than the underlying information sources (e.g. Google News), and pure-play advertising platforms such as Craigslist, which divert revenue from newspaper classified ads.[2]

Despite these and related challenges, much of the commercial information on the web is provided under ad-financed business models. While a growing number of newspaper sites have moved behind paywalls (Chiou and Tucker, 2013), most other commercial information sources rely on advertising. The finance of information production with ad revenue has come under threat with the development of ad blocking technologies. Ad blockers, such as Ad block Plus, are browser extensions that allow users to browse the web without having to see sites' ads. As a

[1] See Giorcelli and Moser (2016).

[2] See, for example. Oberholzer-Gee and Strumpf (2006), Rob and Waldfogel (2006), and Seamans and Zhu (2010), among others. Ad blocking is not the first technology that allows consumption without compensation to rights holders. Digital video recorders such as TiVo allowed users to fast forward through ads, threatening their revenue generating potential. See Wilbur (2008). In 2012 Aereo created a service distributing broadcast television online without paying rights holders transmission fees. ABC sued and won, shuttering Aereo. See http://www.nytimes.com/2014/06/26/business/media/supreme-court-rules-against-aereo-in-broadcasters-challenge.html .

result, many site visitors obtain information without seeing ads, clicking through to advertisers' sites, or otherwise generating revenue for site owners. The use of ad blockers has grown sharply in recent years. By some estimates, roughly a quarter of site visitors used ad blockers by 2016, a magnitude of ad blocking that may produce a serious threat to site revenue and, if revenue losses undermine investment, a possible threat to consumers' access to appealing content.[3] This is the question this paper seeks to explore.[4]

The mechanism we have in mind, as in the traditional literature on the relationship between intellectual property revenue appropriation and supply, is that reduced revenue may undermine a site's ability to invest, which could then manifest itself as a diminished site that is less appealing to potential visitors. Web users might then visit the degraded sites less, reducing the sites' traffic. While getting data on the relevant issues - site traffic and site-specific measures of ad blocker usage - is challenging, we have put together a dataset that allows us to explore this. Using unique, proprietary, and site-specific data on the share of site visitors using ad blockers at a few thousand sites, matched with Alexa traffic data, we explore the impact of ad blocker usage on site traffic between 2013 and 2016.

Aggregate ad blocking has grown rapidly between 2013 and 2016, and during this period, we find that sites with a high proportion of ad blocking visitors experience deterioration in their traffic ranks, relative to sites with fewer ad blocking visitors. In our main analyses, we find that each additional percentage point of site visitors using ad blockers raises (worsens) a site's traffic rank by about 0.6 percent over a 35 month period, with the effects concentrated in sites that are initially lower ranked. We have two ways of dealing with possible concerns about unobserved

[3] http://www.emarketer.com/Article/US-Ad-Blocking-Jump-by-Double-Digits-This-Year/1014111

[4] Anderson and Gans (2011) develop a theoretical model of ad-blocking showing the possibility of a downward spiral in which ad blocking leads websites to carry more ads, further reducing their appeal.

heterogeneity. First, in contrast to our basic result – worsening ranks at sites whose users block ads – those same sites did not experience worsening ranks prior to when we would expect ad blocking to directly impact content investment decisions. Second, when we employ instrumental variable techniques to address bias arising from either measurement error or possible endogeneity concerns, we find consistent results. Given our basic result, if 20% of a website's users employ an ad blocker, then between 2013 and 2016, the site's rank worsens by about 12.5% relative to the counterfactual ranking it would have if none of its users blocked ads. Because we find close to a one-to-one relationship between ranks and traffic, the impact on site traffic is proportionally similar. Effects on revenue are compounded by the fact that ad blocking reduces visits, while also generating less revenue from ad blockers among the remaining visitors. For example, if revenue is proportional to traffic, a twenty percent ad blocking share could reduce revenue by about 30 percent.[5]

The paper proceeds in six sections after the introduction. In Section 1, we present some basic facts about advertising as a means of financing content online, we describe ad blockers and their growth, and we provide a simple theoretical framework for organizing our thinking about the possible effect of ad blocking. The second section describes the data. Section 3 describes our empirical strategy and results on the relationship between ad blocking and site ranks. Section 4 translates the effects on ranks into effects on traffic and revenue. A concluding discussion follows in Section 5.

[5] To see this, note that (1-0.2(0.6))(1-0.2) = 0.704, which is almost 30 percent below unity. We discuss some important details of web advertising that may temper the impact of ad blocking on revenue below.

I. Background

Ad-supported content delivered online is an important mechanism for informing consumers and citizens in the US and elsewhere. In practice, advertising provides far more revenue than access fees (paywalls) for such content. Take, for example, newspapers, which perhaps are the most aggressive users of paywalls and therefore derive a greater share of revenue from consumer payments than do sites in other categories. Notable examples include the *Wall Street Journal* and the *Financial Times*, whose sites are available only to subscribers. Still, the share of digital revenue from users as opposed to advertisers remains small for newspapers generally. In 2013 the Newspaper Association of America reported digital ad revenue at US newspapers totaling $3.49 billion, while digital subscription revenue was roughly $0.75 billion.[6] In other words, even for newspaper websites, ad revenue was over four times their subscription revenue.

As recently as 2012, ad blockers were rarely used. Since then, their use has risen sharply. A joint report by Adobe and PageFair followed the number of active ad blocker installations by measuring periodic downloads of "blocklists," lists used by ad blockers to determine what to block. The information, in Figure 1, shows only 39 million installations worldwide in January 2012. By June 2015, there were 198 million, 45 million of which came from the US. Dividing this number by the number of internet users in the US suggests that as many as 16% of US internet users may use ad blockers. Analogous calculations suggest higher numbers in Europe: 25.5% in Germany, 21.1% in the UK, and 16% in Spain. Comparable statistics from a separate

[6] Multiplying circulation (online and hardcopy) revenue in the same year ($10.87 billion) by the share coming from digital (7 percent), a statistic from 2014, yields and estimated digital subscription revenue of $760 million See http://www.naa.org/Trends-and-Numbers/Newspaper-Revenue/Newspaper-Media-Industry-Revenue-Profile-2013.aspx and http://www.statista.com/statistics/322916/newspaper-circulation-revenue-print-digital-use/

data source, a ComScore panel of monitored internet users, found similar but slightly lower percentages: 24% for Germany, 14% for Spain, 10% for the UK, and 9% for the US. However, the ComScore data also found that users of ad blockers appear to spend more time on the internet, leading to higher percentages for the fraction of website page views which have their ads blocked: 28% in Germany, 16% for Spain, 13% for the UK, and 12% for the U.S. Moreover, its use will likely grow as consumer awareness increases. A recent study found that only 41% of respondents were aware of ad blocking, but among those who were aware, 80% used an ad blocker.[7]

There is substantial controversy surrounding ad blockers. As the Internet Advertising Bureau (IAB) puts it, "As abetted by for-profit technology companies, ad blocking is robbery, plain and simple – an extortionist scheme that exploits consumer disaffection and risks distorting the economics of democratic capitalism."[8] At the same time, many industry participants acknowledge that web advertising has become unwieldy, a problem exacerbated by the fact that obtrusive ads and targeted ads have each been shown to be more effective, (Goldfarb and Tucker 2011a, 2011b, 2015), particularly when consumers' privacy has unwittingly been violated. (Goldfarb and Tucker 2011b, Tucker 2014). As a result, loading a site and its ads consumes substantial bandwidth, which is a particular challenge for mobile users. A 2015 study by the *New York Times* found that advertising and other data blocked by ad blockers on 50 news websites' main mobile homepages accounted for more than half of the sites' data, more than doubling the time it took to load.[9] Moreover, "malvertisements," or malicious ads designed to

[7] http://digiday.com/publishers/survey-80-percent-know-ad blocking-use/
[8] http://www.iab.com/iab-believes-ad blocking-is-wrong/
[9] Aisch, Gregor, Andrews, Wilson, and Keller, Josh, "The Cost of Mobile Ads on 50 News Sites." The New York Times. Oct 1, 2015.

expose vulnerable computers to viruses, have yielded over ten billion impressions in some recent years.[10]

The IAB has itself argued that "ad blocking is a crucial wakeup call to brands and all that serve them about their abuse of consumers' good will. Ad blockers are also exploiting a real vulnerability: the erosion of stimulating consumer experiences online. For this, the marketing-media ecosystem bears real responsibility. IAB research shows ad block use is caused by a general disdain for advertising and concern about the safety of user information."[11]

The rise of ad blocking raises a number of research questions. A first possible question – analogous to the first generation of piracy research following Napster – is simply whether the growth of ad blocking reduces advertising revenue. At first glance this seems like estimating an identity rather than a behavioral relationship since ad blocking disables ads and therefore prevents the generation of ad revenue.[12] But it is possible that some users of ad blocking software would forgo visiting sites if forced to view ads. If so, then the ad blocking visitors would not represent lost revenue. Rather, they would reflect additional, albeit unpaid, users who would not have made site visits absent their ability to block ads.

While revenue is of course important to advertisers and publishers, revenue is not by itself a clear matter of concern for public policy. A more important indicator is whether, because of reduced revenue, the quality and appeal of sites declines. Revenue and quality do not

---

[10] Schlesinger, Jennifer. "Beware of Malicious Ads That Can Harm Computers Without a Click." CNBC. May 20, 2014.

[11] http://www.iab.com/iab-believes-ad blocking-is-wrong/

[12] According to the President of Ad block Plus, the most popular ad blocker, "Before the page is rendered, ad block plus modifies it, strips off the request to the ad service or tracking scripts and injects CSS to repair the site so it doesn't look broken" http://www.computerworld.com/article/2487367/e-commerce/ad blockers--a-solution-or-a-problem-.html. Hence, ad blockers prevent the request for ads – no transaction takes place - implying websites are not paid for blocked ads.

necessarily go hand in hand. In other contexts, most notably the recorded music industry, the collapse of revenue has not brought about a reduction in the quantity or apparent quality of new recorded music. Instead, falling costs have prevented revenue reduction from undermining the production of content.[13]

These nuanced considerations suggest a second research question – addressed here – of whether ad blocking leads ad-supported websites to degrade their sites in response to declining revenue. That is, do they reduce their expenditures on content and produce less, or lower-quality, content? This question is inherently difficult to study, at least directly, for a few reasons. First, we cannot observe sites' expenditures on content. Second, while we could conceivably measure the amount of content, measuring its quality is difficult to do directly. Some media products, such as music and movies, are evaluated by large cadres of professional critics, and their evaluations can be useful for assessing the quality of those products. Nothing like this exists systematically across web sites, however.

Instead, we infer "product quality" from the traffic at a website, an indirect method frequently employed by industrial economics in demand estimation. It is reasoned that consumers are attracted to a product by its various horizontal and vertical attributes that manifest themselves for consumers in the "mean utility" of the products. The econometrician infers this mean utility, in turn, from the products' market shares, along with their prices. If a product's market share falls, then all else equal we can infer that its mean utility (i.e. quality) has fallen relative to alternatives. Hence, if websites that experience more ad blocking eventually have

[13] See, for example, Waldfogel (2012).

larger reductions in traffic, compared to sites experiencing less ad blocking, we can infer that their quality has fallen.

a. Theoretical Framework: Short vs. Long Run

A simple model is helpful for analyzing the possible effects of ad blocking technology on the market for ad-supported online content. Users, denoted by i, differ in the value that they attach to the experience of visiting a site, and the demand curve shows the distribution of these valuations across users. Prior to the advent of ad blocking, users "pay" for site content in the sense of enduring the nuisance cost of advertising, along with the users' perceptions of the cost of lost privacy. Figure 2 depicts the situation. Users attaching a valuation (v_i) to the site that exceeds the nuisance cost of the ads (n) visit the site. As a result, the site attracts q_0 visitors. Ads generate revenue of p per user visit, so the site generates revenue of pq_0.

The advent of ad blocking changes the "price" of visiting the site, and the impact of ad blocking can differ between the long and the short run. In the short run, users can choose, by installing ad blockers, to face a lower price of visiting the site. In the longer run, it is possible that the site quality will change, prompting a shift in the demand curve. This shift could, potentially, more than offset the reduced nuisance cost.

We begin with the short-run impact. Users employing ad blockers, freed of the nuisance cost, now face a zero cost of visiting the site. This has different effects depending on whether the user's valuation exceeds the former nuisance cost "price" of visiting. Users whose valuations exceeded the nuisance cost continue to visit. Users whose valuations fall short of the nuisance cost will now visit the site, if they have installed an ad blocker. Suppose a share α of users installs ad blockers and that those visitors generate no revenue for the site. For simplicity, assume that the ad blocking visitors are drawn uniformly across the site valuation distribution. Then the number of visitors rises from q_0 to $q_0 + \alpha(q_1\text{-}q_0)$. Revenue declines from pq_0 to $(1\text{-}\alpha)pq_0$. Thus, the short run effect of ad blocking on the market will be an increased quantity of visitors, along with a reduction in revenue (if at least some of the ad blocking visitors were previously visitors).

In the longer run, which we define as the period after which the site can adjust its investment in content and therefore site quality, content investments may decline, reducing the resultant desirability of the site, and giving rise to the leftward shift in the demand curve. The longer-term impact of ad blocking on site visits is now ambiguous. To see this, consider some special cases. First, suppose that all users install ad blockers and that site quality declines, giving rise to the left-shifted demand curve in Figure 3. Then the visit quantity will rise, in the long run, from q_0 to q_2. Of course, if demand had shifted farther to the left, so that the new demand curve hits the quantity axis at a quantity below q_0, then our conclusion would have been reversed: if all users installed ad blockers, then their long run impact would be to reduce the number of site visitors.

The short run and long run comparison is helpful for guiding our analysis. Absent any quality changes, ad blocking leads to an unambiguous increase in traffic, by lowering the nuisance cost of visiting the site. Thus we would expect non-negative (and possibly positive) impacts of ad blocking on website traffic in the short run. Moreover, if traffic does decrease over a longer-term horizon at sites experiencing more ad blocking, we can infer the demand curve has shifted to the left, i.e. site appeal at those sites has declined. This is a simplified analog of the downward circulation spiral modelled in Anderson and Gans (2011).

b. The Complication of Direct Ad Sales

The foregoing analysis embodies the assumption that site revenue is proportional to the number of non-ad blocking page impressions, and this is a reasonable approach for "programmatic" ads, such as those sold in ad exchanges. But some sites also have an in-house sales team selling "direct" ads, whose contracts promise a particular number of impressions. Direct ads typically generate five times as much revenue per impression than do programmatic ads.[14] The use of direct of ads, generally in conjunction with programmatic ads, can mute the effect of ad blocking.

To see this, consider the following setup. A site has total traffic of T per period. Of these T impressions, the site has sold D impressions directly at a price of p_d per impression. The remaining impressions $(T - D)$ are sold at one fifth the price, $\frac{p_d}{5}$ per impression. Hence, site ad revenue is $p_d D + \frac{p_d}{5}(T - D)$, or more simply, $revenue = \frac{p_d}{5}(4D + T)$. Now introduce ad blocking. If one more impression is blocked, so that T falls by 1, then as long as the total number of impressions that are not blocked exceeds the direct sales quantity D, the direct sales revenue is unaffected. Instead, only programmatic ad revenue is affected. Overall, revenue falls by $\frac{p_d}{5}$, to $\frac{p_d}{5}(4D + T - 1)$. However, when the number unblocked impressions falls below D, additional blocked impressions reduce revenue by p_d. Hence, the effect of ad blocking on revenue at sites with direct ad sales has a kink: each lost impression reduces revenue by $\frac{p_d}{5}$ until

[14] Based on information by an industry source that preferred to remain anonymous. Less specific but publicly available information is available at http://buyercloud.rubiconproject.com/content/101/programmatic-inventory/

the number of blocked impressions reaches $(T - D)$; thereafter each lost impression reduces revenue by p_d, i.e. at five times the rate. See Figure 4.

Note that an additional blocked impression at a site without direct ads reduces revenue by $\frac{p_d}{5}$, or by the same absolute amount as at a site that also uses direct ads. The proportionate reductions differ, however. If a small share π of impressions are blocked, so only programmatic ads are blocked, then revenue falls by $\pi T \frac{p_d}{5}$. The percentage reduction in revenue is then $\frac{\pi T p_d/5}{(p_d/5)[4D + T]}$, or $\frac{T\pi}{4D+T}$. For sites not selling direct ads, this simplifies to $T\pi/T$, so that if a share π of impressions is blocked, then revenue falls by 100π percent. If a site has directly-sold ads, the percent reduction in revenue is smaller, at least initially. Eventually, however, if ad blocking grows enough, the marginal impact of ad-blocking will become far more severe.

Selling direct ads requires a direct sales force; only high-traffic sites have sufficient revenue to engage in direct sales. Hence, we will look below for evidence of smaller impacts of ad blocking on site quality for initially-larger sites.

II. Data

Because the possible effects of ad blocking on traffic might take time to operate, we would ideally have data on both the number of visitors and the share using ad blockers, by time and website, over a period beginning prior to the appearance of ad blockers. We would then ask whether the sites experiencing larger growth in ad blocking experienced declines in quality, as inferred from the long run tendency for consumers to visit the sites. Our actual data deviate from the ideal setup, while still offering some promise for measuring the impact of ad blocking on site success. We have a fairly long (three year) panel on site traffic, but we have only intermittent data on the share of site visitors blocking ads at particular sites.

We have site traffic data from Alexa, a company that provides website traffic ranks, for April 2013 through June 2016. The data cover roughly 34 million websites, and the ranks are calculated based on the last 3 months of page views.[15] [16] Our data on the share of site visitors blocking ads are from PageFair (https://pagefair.com/), a company that provides software that site owners can use to monitor the share of their site visitors using ad blockers. PageFair's site-specific data on ad blocking and traffic are available intermittently between 2013 and 2016, for different time spans depending on the site. While it would be desirable to have a website-level panel with information on traffic and the extent of ad blocking by their visitors, the available data do not consistently cover a long enough span of time corresponding to the growth in ad blocking, because many website owners elect to use the software on a temporary basis. The median duration of ad blocking that we observe at our sites is 16.7 weeks (see Figure 5). The site monitoring episodes begin in different months between 2013 and 2016, but the majority of

[15] Engineers at PageFair found that Alexa's toolbar extension, which measures site traffic, does in fact count traffic from users of ad blockers, even when using an aggressive ad blocker, uBlock Origin. Moreover, we ran several (unreported) statistical tests relating Alexa's ranks to true total traffic levels measure directly by PageFair, during periods of overlap. We found no evidence that websites with higher (or increasing) rates of ad blocking had worse ranks than expected from their true, overall traffic levels measured by PageFair (including traffic from users of ad blockers). Additionally, we found the Alexa traffic rank of Bild.de, the website of a major tabloid newspaper in Germany, worsened substantially after it banned ad blockers from visiting their site. See Figure 12. This worsening in ranking would only be expected if users of ad blockers were counted in Alexa's traffic data.

[16] Alexa's traffic ranks are computed using two sources of traffic data, traffic meters installed at sites that participate in Alexa's monitoring and individuals' web browsing data collected from Alexa's toolbar extension and other undisclosed toolbar extensions. See https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined-. From Alexa's summary pages for individual websites (e.g. http://www.alexa.com/siteinfo/cbs.com), we were able to deduce that the latter source (toolbar-measured traffic) is used exclusively to compute ranks for 99% of the websites in our data; the text "Is this your site? Certify your metrics" appears only for websites that have not installed Alexa's monitoring.

PageFair's monitoring episodes begin after June 2015. We match Alexa traffic data with 2,574 sites that appear in PageFair.

Because the PageFair panel data on site-specific ad blocking cover only short time periods, we effectively have only a cross-sectional measure of ad blocking for each site, mostly for a period starting after mid-2015. We use the data to measure each site's ultimate shares of ad blocking visitors (see below). Putting some details of implementation aside for the moment, if we have each site's ultimate ad blocking share, along with an assumption that ad blocking was negligible prior to 2013, then our measure of each site's ultimate share also provides an estimate of the site's *change* in the share of its visitors using an ad blocker.

We explore a few different specific methods for calculating the site-specific eventual extent of ad blocking. First, we take the simple approach of dividing the total number of ad blocking visitors at a site by their total number of visitors, over the time period monitored by PageFair. The ratio of ad blocking visitors to total visitors is then our measure of the propensity for the site's visitors to use ad blockers. Define this ratio as P_i^1. Figure 6 shows the density of P_i^1. The median and mean ad blocked shares are 10.1 and 14.6 percent, respectively. The inter-quartile range runs from 5.2 to 18.4 percent.

Our second approach accommodates the disparate timing of different sites' monitoring. We begin by calculating a monthly site-specific share of visitors employing ad blockers, s_{it}, where *i* indexes site and *t* indexes month. We then regress the log odds ratio of this measure on site and time fixed effects:

$$\ln\left(\frac{s_{it}}{1-s_{it}}\right) = \delta_i + \phi_t + \epsilon_{it}.$$

Based on the regression, we can predict the percent of ads blocked at any arbitrary time t, even if the website was not present in PageFair's dataset that period. We define P_i^2 for the final period in the PageFair dataset Feb 2016, as:

$$P_i^2 = 100 * \frac{\exp(\delta_i + \phi_{Feb,2016})}{(1 + \exp(\delta_i + \phi_{Feb,2016}))}.$$

This is an estimate of the eventual ad block rate. As it turns out, our two measures are highly correlated.[17] P_i^1 has the virtue of greater simplicity, so it is our preferred approach to measuring eventual ad blocking shares. We verified that all reported results that we obtain using P_i^1 also arise when using P_i^2.

What sort of sites appear in the PageFair data? We have 2,574 sites in PageFair for which we have an Alexa rank in June of 2013. Of these sites, the median site is ranked 210,000 in Alexa. The inter-quartile range runs between Alexa ranks of 38,000 and 1.4 million. Figure 7 shows the density of log Alexa ranks for June 2013. One notable feature of the figure is the truncation of the distribution at e^{17}, which corresponds to the 34 million cutoff for inclusion in Alexa.

What else do we know about the PageFair sites? First, we have the domain suffixes, shown in Table 1. Most of the websites have "com" as their suffix, but a substantial number have other suffixes, including those that indicate different country locations. For example, 158 have "uk," 76 have "de," and so on. Second, for most websites, we have the percent of visitors from each country, for up to the top five countries at each site (reported by Alexa). The 25 most

[17] Most websites were observed by PageFair in later years, after ad blocking's rise had already commenced. Only about a tenth of the sites joined before 2014, and only a third before 2015.

frequently observed countries are shown in Table 2. One thousand websites with low ranks at time of data collection (July 2016) had fewer than five countries listed. For nearly 500 of these, no information about visitors' geography was available.

Summing up, we ultimately have monthly measures of Alex traffic ranks for PageFair sites between June 2013 and April 2016. In addition, we have measures of the eventual share of visitors blocking ads at their sites, mostly drawn from the period after early 2015. Lastly, we have some information on the geographic location of sites (from suffixes) and origins of visitors.

III. Empirical Strategy and Results

1. Empirical Strategy

We employ three separate empirical strategies for measuring the causal impact of ad blocking on site traffic. Our first approach regresses the change in site rank, 2013-2016, on our measure of the change in the site's share of ad blocking users. This would support reasonable causal inference if the variation in eventual ad blocking varied exogenously across sites. This approach would be undermined, for example, if users made site-specific decisions about whether to block ads; in that case, a site with falling quality might effectively invite increases in ad blocking. But we understand ad blocking to be a user, and not a site-specific, decision. Hence websites are exposed to different rates of ad blocking based on their user composition prior to ad blockers' growth, irrespective of their own traffic trends. For example, we cited evidence above that the use of ad blockers varies across countries, and hence sites catering to different geographic regions face disparate exposure.

Despite its plausibility, our basic approach could be vulnerable to a concern about unobserved heterogeneity, in particular that the sites with high eventual ad blocking shares were already declining, or destined to decline, prior to the advent of ad blocking. This leads us to two approaches for dealing with unobserved heterogeneity in site rank trends that may be correlated with ad blocking, a longitudinal approach and instrumental variables.

First, we verify that the relationship between rank degradation and eventual ad blocking rates appears only when we would expect it to, and not before. Recall that use of ad blockers was negligible prior to 2013, then started growing quickly. Since our proposed mechanism is expected to take some time to operate, we would not expect ad blocking to have a large direct impact on site quality during 2013. Hence we can test for pre-existing trends by comparing regressions of rank changes *after* 2013 on eventual ad blocking with the relationship between rank changes *during* 2013 and eventual ad blocking. This allows us to see whether the effect we attribute to ad blocking arises prior to the time when we would expect widespread ad blocking to have a direct impact on content investment decisions.

While the longitudinal approach deals with possible pre-existing rank trends that are correlated with eventual ad blocking, it leaves open the possibility of an unobserved factor that raises ad blocking and worsens page traffic coincident with the widespread adoption of ad blocking after 2013. We explore this with an IV strategy based on the geographic diffusion of ad blockers.

2. Graphical Evidence

Before turning to regression-based approaches, it is instructive to simply ask how the change in website ranks, 2013-2016, relates to the extent of their visitors' ad blocking. Figure 8 does this with a graph of the smoothed changes in log ranks against the sites' ad blocking visitor shares.[18] As the percent of site users blocking ads runs from 10 to 30 percent, the percent change in site rank rises from 0 to about 25 percent. We perform somewhat more sophisticated tests below, but Figure 8 presents the basic relationship documented in the paper, that sites with proportionally more ad blocking users experience greater worsening in their ranks between 2013 and 2016.

The damaging long-term effect of ad blocking on web traffic ranks apparently operates when we would expect it to operate. But does it fail to operate when we would expect it not to, prior to the widespread use of ad blocking? As Figure 1 shows, usage of ad blockers is low prior to 2013, then rises quickly. Our proposed mechanism should take some time to operate, so if we look at the relationship between the ultimate ad blocking share and rank degradation during 2013, we should either see a positive effect of ad blocking on web traffic (negative impact on rank) – the short-run effects described in the theory section – or less of a negative effect than we see in the long run.

Figure 9 reproduces Figure 8, with two separate smoothed representations of the change in log ranks (on the vertical axis). The blue line shows the change between December 2013 and April 2016. The red dashed line shows the change between June 2013 and December 2013. The rank worsening associated with ad blocking appears clearly for the latter period, but it does not

[18] The figure shows a smooth lowess regression of change in log rank on the site specific measure of ad blocking between June 2013 and April 2016, for approximately the bottom 95% of ad blocking (websites with less than 40% ad blocking). Websites with censored Alexa ranks were assume to have a rank of 34 million.

appear for the second half of 2013. Hence, the relationship between ad blocking and traffic that we attribute to ad blocking does not appear to reflect pre-existing trends.

3. Regression Evidence

a. Direct cross-sectional evidence

Our basic cross-sectional empirical approach is to regress the change in log rank at each site on the site's measure of the share of visitors blocking ads:

$$\Delta \ln(rank_i) = \alpha + \gamma \Delta P_i + \varepsilon_i \qquad (1)$$

In this regression $rank_i$ is the Alexa rank for site i, ΔP_i is the (change in the) share of visitors to site i blocking ads, and ε_i is an error term. Under the empirically reasonable assumption that ad blocking was negligible as of 2013, the site specific measure P_i^1 is therefore also a measure of the change in the tendency of visitors to use ad blocking between 2013 and 2016, i.e. $\Delta P_i = P_i^1$.

Some sites have missing ranks in the final period, either because their traffic was too low to be included in Alexa's rankings, or, in some cases, because they had shut down. Hence, for these sites, we do not observe the dependent variable in equation (1), the change in log rank. However, because a missing rank means that the post rank exceeds 34 million, if the initial rank is, say, 5,000, then we know that the change in log rank is at least ln(34,000,000) – ln(5,000) or, more generally, that the change in log rank ≥ln(34,000,000) – ln(initial rank for site *i*). Thus, we can include these observations, on sites with missing rank, using censored regression. In particular, we assume normality. Then these observations enter the likelihood function reflecting the probability that $\Delta \ln(rank_i) \geq ln(34{,}000{,}000) - ln(initial\ rank\ for\ site\ i)$. Non-missing observations enter the likelihood as the probability that

$\Delta \ln(rank_i)$=$ln(final\ rank\ for\ site\ i) - ln(initial\ rank\ for\ site\ i)$.

Table 3 reports various regression specifications for the full period (June 2013 – April 2016) on our basic measure of ΔP_i, i.e. P_i^1. Column (1), using the OLS approach that treats a site with missing rank as one with rank of 34,000,000, shows that an increase in the share of visitors blocking ads from 0 to 1 (100%) raises the change in log rank by about 0.54, or, said another way, an increase (i.e. worsening) in final rank by roughly 54%. Column (2), which uses the interval regression approach, yields a higher coefficient of roughly 0.63.

Although we report parsimonious specifications in Table 3, one might prefer regressions with controls out of concern that some potentially observable factor is correlated with both ad blocking and traffic trends. For example, suppose that site suffix, a rough proxy for country, is correlated with declining interest in the Internet. Then sites appealing to members of the corresponding group could have declining traffic, and the same users losing interest in the web might be more prone to use ad blockers (because they are no longer willing to suffer through ads to get to content). To address these concerns, we also explore specifications using website suffix dummies as a control.[19] We obtain similar results. Censored regression with these controls gives a coefficient of 0.63, with a standard error of 0.20.

Based on the analysis in Section 2b, we might expect ad blocking to have a dampened effect at websites that sell direct ads, which we understand to be more common at high-traffic websites. We explore this possibility by running a censored regression of log rank change over the period on a website's ad blocking rate, a "high-traffic" indicator variable equaling one for

[19] We assign suffixes that appear fewer than 25 times to an "other" group.

websites in the top quintile of initial rank, and their interaction. The regression specification, estimates, and standard are shown below.

$$\underset{}{\Delta \ln(rank_i)} = \underset{(0.05)}{-0.17} + \underset{(0.23)}{0.79}\ P_i + \underset{(0.10)}{0.49} * I(hi\ traf_i) - \underset{(0.46)}{0.91} * P_i\, I(hi\ traf_i) + \epsilon_i$$

The interaction coefficient is significant. Hence, we can conclude ad blocking had a significantly smaller impact at high-traffic websites. Due to large standard error for the interaction term, the overall impact of ad blocking on premium websites' rankings is statistically indistinguishable from zero. Repeating this analysis, but instead defining premium websites as those in the top 10% of initial rank, yields similar results. These results are consistent with the idea that high traffic sites, which are likely to derive revenue from direct sales, have so far experienced smaller effects of ad blocking.

Finally, because some websites have relatively few visits in the PageFair data, our ad blocking variable *P* may be measured with error, biasing coefficient magnitudes towards zero. Table 4 explores this possibility by re-running the basic OLS and censored regressions (the first two columns of Table 3) dropping observations which compute ad block rates from few visits. When we drop these observations, the coefficient estimates rise somewhat, from 0.55 to 0.82 with OLS and from 0.63 to 0.91 with the censored approach.

b. Pre-existing trends

We can supplement the evidence in Figure 9 with regression evidence that the relationship between ad blocking share and rank worsening does not hold before we would expect it to. If we have multiple periods of observations on the change in site ranks, we can estimate the following equation:

$$\Delta \ln(rank_{it}) = \alpha_t + \gamma_t \delta_t \Delta P_i + \varepsilon_{it} \qquad (2)$$

where δ_t is an indicator that takes the value of 1 in the corresponding period. The coefficient γ_1, corresponding to the first time interval, shows the relationship between eventual ad blocking and the rank changes in the period before we would expect ad blocking to reduce investments in content. We have data on a 35-month period running from June 2013 to April 2016, which we divide into five equal-size periods of seven months, denoted *t*=1,…,5.

Table 5 reports estimates of (2) using both OLS and the censored approach, in columns (1)-(4). In all specifications γ_1 (the coefficient for the first time period) is negative and significant, indicating that the sites with higher eventual ad block penetration experience improvements in traffic during the early period. This is consistent with our short run theoretical prediction that ad blocking would improve traffic prior to its possible effects on the site appeal. We also see that the relationship between rank changes and eventual ad blocking changes for periods after the first. Columns (2) and (4) show the evolution of the γ_t coefficients, and they go from negative and significant in the initial period to positive and significant in subsequent periods. Columns (1) and (3), which constrain γ_t to be the same in all periods beyond the first, allow for a direct test of whether γ_t differs from γ_1 for t>1. We reject the equality in both linear and censored specifications.

We can take this approach one step further to explicitly eliminate the time-constant unobserved heterogeneity. Rewrite equation (2) as

$$\Delta \ln(rank_{it}) = \alpha_t + \gamma_t \delta_t \Delta P_i + \mu_i + \omega_{it} \qquad (3)$$

where the error ε_{it} has two explicit components: $\mu_i + \omega_{it}$. The term μ_i represents time-constant unobservable determinants of the change in ranks (including a pre-existing trend in traffic that may be correlated with the tendency for site users to employ ad blockers), while the term ω_{it} is a site-and-time idiosyncratic error. The inclusion of a site fixed effect in (3) purges time-constant unobserved heterogeneity but precludes the estimation of α_1 and γ_1, the coefficients showing the relationship between eventual ad blocking and rank changes in the period prior to widespread diffusion of ad blockers. Instead, we can estimate γ_t, for t=2,…,5, which shows how the relationship between rank changes and eventual ad blocking *evolves* with the diffusion of ad blocking. As columns (5) and (6) in Table 5 show, the γ_t coefficients for the periods after the initial period are significantly larger than for the initial period. These results are consistent with the idea that the diffusion of ad blocking reduces site traffic.

c. Instrumental Variables Approaches

The remaining endogeneity threat, referring to equation (3), is that the contemporaneous error ω_{it} is itself endogenous. This would arise, for example, if site i grew unappealing in a period t of widespread ad blocker adoption, prompting both (a) rank degradation and (b) user unwillingness to view a site without an ad blocker enabled. Put another way, a third unobserved factor could be driving both the ad block share P and the site's log rank. We can address this somewhat remote possibility via instrumental variables if we can find a source of variation in eventual ad blocking that does not affect site ranks except via its effect on ad blocking.

The geographic distribution of visitors to each site provides an appealing instrument for the extent of ad blocking, since the fraction of internet users eventually gaining the capability of blocking ads (i.e. installing blocking software) varies substantially across countries, and websites catered to different territories prior to the growth in ad blocking. Hence, sites with different visitor origin shares are subjected to different extents of ad blocking for reasons unrelated to site appeal. Instrumental variables address this concern, as well as allow us a possible remedy for measurement error in the extent of ad blocking, *P*.

We pursue a two approaches to the geographic IV approach. First, we use the website suffix, which indicates the country in which the domain name is registered. For example, a site ending in ".uk" is registered in Great Britain. Because many suffixes are infrequently observed (see Table 1), we assign all suffixes with less than 25 observations to the "other" category, which leaves us with twelve suffix dummies. A regression of *P* on this set of site suffix dummies yields and F-statistic of 9.85. The resulting second-stage regression is shown in column (3) of Table 3, and the coefficient on *P* is similar to the magnitude when using OLS, although the standard error is larger (s.e.=0.92).

Our second geographic approach to instrumenting for *P* uses the percent of visitors from each of up to five countries, reported by Alexa. Alexa typically reports the top five countries generating visitors to each website, but for less popular websites they report fewer. For nearly 500 websites, none are reported.[20] We include this latter set of instruments in two ways. First we use them directly, dropping websites for which Alexa has no information on visitor origins, and we restrict attention to the visitor countries with at least 25 observations. The first stage is thus a regression of *P* on 43 variables such as the percent of site visitors from the US. The resulting F-statistic for the first stage is 5.70, and the second stage regression is reported in column (4) of Table 3. The coefficient of interest increases from 0.54 under OLS to 0.86, with a

[20] For 460 websites, no country data were available. For an additional 36 websites, only data on visitors from obscure countries (countries with fewer than 25 observations) were available. Both sets, 496 in total, are considered missing.

standard error of 0.61.

Recall, however, that we are completely missing the visitor geography data for about 20 percent of sites with the lowest site traffic when the data were collected, in June 2016 – the end of the time period. Hence, we are systematically dropping sites that substantially deteriorated. The remaining sample hence only includes the heavily ad blocked websites which were fortunate enough to remain viable despite the impact of ad blocking. Such selection could yield estimates which understate the true effect. To explore this, we impute traffic origin shares for the websites with no origin data reported, by setting the missing origin country visitor percentages equal to the averages at websites with the same suffix.[21] Using this imputation, the coefficient rises to 1.43. The resulting F-statistic is 3.02. Imputation complicates standard error calculation. We use a bootstrapping approach, taking 1,000 draws from the observed distributions of the percentage measures, conditional on each suffix.[22] The resulting standard error is 0.49.

In summary, the IV approaches give larger but less precise estimates of the impact of ad blocking on traffic rankings. For the sake of conservatism we employ the smaller but more precisely estimated OLS estimates as our baseline approach.

d. Additional Robustness: Regression to the Mean

Our sampling scheme involves watching ranks evolve, 2013-2016, for a sample of sites that participate in PageFair's monitoring program. While it's a bit hard to generalize about the PageFair sample, we can say that most of these are firms that opt into this program in 2015 or

[21] To remain consistent with the non-missing observations, we only include the five countries with largest imputed visitor percentages at each website.

[22] The bootstrapping procedure repeatedly resamples websites (with replacement). For each drawn sample, we redo the imputation of origin visitor shares for the websites in the sample missing such data, using the average visitor shares at websites with the same suffix in the drawn sample, and then re-run the IV regression. The reported standard errors are the standard deviations of the coefficients from these 1,000 redrawn and re-imputed samples.

later. That is, the sites are healthy enough nearly two years after the start of the rank monitoring to seek to participate in the program. Given that the PageFair sites may be sites whose traffic is systematically rising or falling relative to other sites, it is possible that their ranks would change over time in the absence of ad blocking.

We explore this in Table 6 by relating the change in log ranks to the initial rank. The first two columns regress the change in log ranks on the ad blocking share (P), along with the initial Alexa rank and, in column (2), its log. Both are statistically significant predictors of the change in log rank. The third column includes dummies for the initial rank decile. In this flexible specification, the coefficient of interest (on P) is 0.65, very close to the value (0.63) of the coefficient in the analogous regression in column (2) of Table 3. We conclude that our estimate is robust to the inclusion of controls for the initial rank.

IV. Effects on Site Traffic and Revenue

While the relationship between ad blocking and a website's rank shows an important directional impact of ad blocking on the relative appeal of a site, it is difficult to interpret the economic impact without more information. Using cross-sectional data on site ranks along with measures of traffic, we can translate the changes in rank we document above into changes in traffic. The calculation is different depending on whether a site sells direct ads. We begin with the case of a site with only programmatic ads, so that are revenue is proportional to non-ad-blocked impressions.

For the calculation we need a traffic quantity measure along with a rank, for a range of websites. In addition to the historical traffic ranks employed above, Alexa also reports a current measure of usage, the percent of global page views. Alexa maintains a list of the top million sites based on the most recent month's traffic.[23] We drew, from this list, the top 2000 websites along with every thousandth site, by rank, from 2,000 to 1,000,000. For each, we then found the three-month rank and page view measure. The resulting three-month ranks ranged from 1 to 2,166,542. Figure 10 summarizes the relationship, and Table 7 presents a regression of log page views on log ranks. The coefficient on log rank is -1.06 and fairly precisely estimated.

We can now translate our estimates into effects of ad blocking on traffic, i.e. quantity. Above we showed that $\frac{dln(rank)}{dP} \approx 0.6327$. Here, we see that $\frac{dln(traffic)}{dln(rank)} \approx -1.06$. Hence, our estimate of the impact of P on site traffic is the product of these two estimates, or about -0.67. That is, an increase in ad blocking from 0 to 100 percent would decrease site visits by 67 percent. Given the mean ad blocking share in the sample is about 12 percent, ad blocking at the average site reduces overall traffic by about 8%.[24]

If, as we have argued, ad blocking reduces the number of visitors, then the impact of ad blocking on revenue has two parts. First, holding constant the number of visitors, ad blocking reduces revenue because some of the visitors – those blocking ads – do not generate revenue. Second, ad blocking appears to reduce the number of site visitors. Hence, ad blocking has a dual impact on revenue.

If share P of visitors block ads, then according to our estimates, the site will experience a $0.67 * P$ share reduction in its number of visitors. Equivalently, the number of visitors is reduced to share equal to (1- 0.67*P) of its initial visitors count. Moreover, revenue is

[23] https://support.alexa.com/hc/en-us/articles/200449834-Does-Alexa-have-a-list-of-its-top-ranked-websites-

[24] Any impact on content quality may be compounded by declining visibility. As site quality declines, traffic will decline, causing their search engine rank to worsen, reducing traffic even further.

generated only from the share of visitors who do not block ads, i.e. share $(1 - P)$ of visitors. Hence, the number of visitors continuing to generate ad revenue for the site at the end of our 35 month period equals $(1-P)(1- 0.67P)$ times the initial number of visitors. If we assume all traffic yields the same expected ad revenue, then revenues are likewise $(1-P)(1- 0.67P)$ times the initial revenue. Hence, when P=0.12, the average, revenue declines by nearly 20 percent. See Figure 11.

If a site has direct ads, then while the absolute loss in revenue will be the same as at sites with only programmatic ads, the proportional impact on revenue will be substantially smaller as long as the number of non-ad-blocked impression exceeds their direct sales quantity. Thereafter, increases in ad blocking would have larger – and more than proportional – impacts on revenue.

V. Concluding Discussion

The Internet has delivered a variety of positive and negative shocks to appropriability of the value of products in content industries. Just as digitization has delivered strong negative shocks to revenue in recorded music and newspaper industries, digitization – in the form of ad blocking - is also likely having substantial effects on the revenue for a variety of sites delivering ad-supported content online. Digitization-induced revenue reduction does not in all circumstances undermine investment incentives. In the cases of music, movies, and books, falling costs of bringing products to market seem to have more than offset any negative impacts of revenue reduction on creative incentives.[25] But what we have presented here is not simply

[25] See, for example, Waldfogel (2012).

evidence of reduced revenue. Indeed, we have no direct evidence on revenue. Rather, our results show that sites with more users who block ads experience reductions in traffic which we presume arise from the sites' loss of revenue. Our interpretation of the result is that revenue reductions undermine investment which, in turn, compromises site quality, making consumers less interested in visiting in the first place. Our interpretation is lent some credibility by industry participants such as ArsTechnica, a popular site among technology enthusiasts. The site's owners report that:

> Most sites, at least sites the size of ours, are paid on a per view basis. If you have an ad blocker running, and you load 10 pages on the site, you consume resources from us (bandwidth being only one of them), but provide us with no revenue. Because we are a technology site, we have a very large base of ad blockers. Imagine running a restaurant where 40% of the people who came and ate didn't pay. In a way, that's what ad blocking is doing to us. Just like a restaurant, we have to pay to staff, we have to pay for resources, and we have to pay when people consume those resources. ... It [Ad blocking] can result in people losing their jobs, it can result in less content on any given site, and it definitely can affect the quality of content.[26]

Our results provide quantitative evidence consistent with the concerns raised. Ad blocking appears to have a large and meaningful impact on the quality of content created.

In the period studied, the impact of ad blocking appears most pronounced for websites that were not initially among the most highly ranked. This is consistent with the idea that higher-traffic sites rely at least partly on direct ad revenue. Even if they are somewhat shielded from effects of ad blocking, major sites do not seem immune to concern about ad blocking; and some site publishers have taken action to reduce the impact of ad blocking. Facebook, the 3rd ranked website globally, began implementing technology which disguised their ads in August 2016, allowing the ads to bypass ad blockers. Public statements suggested these actions may have

[26] Fisher, Ken. "Why Ad Blocking is Devastating To the Sites You Love." Ars Technica, Mar 6, 2010.

raised desktop ad revenues by nearly 10%.[27] As of this writing, Forbes bans visits by ad blocking users, a move which may be designed to stem revenue losses now or in a future of more widespread ad block usage.

One potential solution to ad blocking by site visitors is to deny access to visitors who are using ad blockers. A user running an ad blocker has two choices in the face of a site that forbids ad blocking visitors. The user can disable his or her ad blocker, or the user can leave the ad blocker running and forgo visiting the site. The relative appeal of these two decisions presumably depends on the distinctiveness of site content.

The experience of the Bild.de site in 2015 is instructive. According to Alexa, Bild was the 15th-ranked site in Germany and the 436th ranked site in the world (as of July 12, 2016). Late in 2015, Bild began to prevent ad blocker users from visiting their site (see Figure 12). Their traffic rank quickly worsened over a three month period, rising 10 percent relative to comparable sites, before leveling off.[28] Since Alexa ranks include the last three months of traffic, this is the pattern we would expect if some users substituted toward other sites, perhaps to those not preventing ad blocking visitors. Similarly, Forbes.com, the 220th most popular website globally (as of September 14, 2016), ran an analogous experiment on a subset of users, finding 58% of ad blocker users decided to leave the site rather than disengage their ad blocker.[29] These considerations have led some publishers to contemplate coordination in their efforts to deny access to ad blocking users. The Financial Times reports that "90 percent of Sweden's

---

[27] http://www.wsj.com/articles/5-things-marketers-should-note-from-facebooks-third-quarter-earnings-1478128415
[28] Alexa defines "related sites" using the following data: keyword traffic, audience demographics, frequently of co-citation on other webpages, co-ownership, domain name similarity, and category overlap. Comparable sites for Bild.de were: faz.de, focus.de, spiegel.de, stern.de, taz.de, transfermarkt.de, web.de, welt.de, xing.com, and zeit.de. See http://www.alexa.com/siteinfo/bild.de, accessed June 2016
[29] Dvorkin, Lewis. "Inside Forbes: From 'Original Sin' to Ad Blockers – And What the Future Holds." Forbes, Jan 5, 2016.

publishers (about 20) plan to collectively block the ad blockers during the month of August" of 2016.[30] As of September, 2016, the experiment had not materialized.

If ad blocking undermines ad supported business models online, market failure – in the sense of sites not being able to generate revenue to cover their costs - is not inevitable. Some newspapers have successfully turned to paywalls so that they can generate revenue directly from users. This solution has two problems. First, paywalls have proven somewhat difficult to implement, except for the most distinctive products. Just as users can substitute one site that does not block ad blockers for another that does, they can substitute a free site for a paid site.[31] While not a perfect substitute for a local news site, the free cnn.com provides an attractive source of news that probably limits the ability of other news sites to charge. Second, even if sites succeed in instituting paywalls, doing so may introduce another market failure. Information is a zero marginal cost product, so a positive price will prevent some socially valuable consumption (in which consumers' valuations exceed cost).

Furthermore, users may turn to ad blocking because websites, when choosing the extent and invasiveness of ads, do not internalize an externality imposed on other websites. As Ken Fisher, the Founder of Ars Technica, explained, "the majority of people blocking ads on our site were doing it because other sites were irritating them … It's the worst players in the web publishing world that's driving this."[32]

One last point is in order. If our results are correct, they point to an interesting distinction between behavior that is beneficial for an individual consumer as opposed to what is beneficial

[30] Cookson, Robert. "News Media Move to Ban Ad Blockers from Websites." Financial Times, Jul 6, 2016. It is not clear whether this coordinated approach to ad blocking would be legal under US antitrust law.
[31] Chiou and Tucker, 2013, found that traffic declined by 51% after media websites instituted paywalls.
[32] Murphy, Kate. "The Ad Block Wars." The New York Times. Feb 20, 2016.

for consumers as a whole. While an individual web user might enhance his or her experience by blocking ads, when many users do so, the effects on revenue and investment can undermine the quality of ad supported content, leaving all consumers worse off. Unless technology offers a solution, the possibility of an ad blocker's dilemma raises a question of whether regulation – or even coordination among publishers – might enhance the welfare of site visitors along with the sellers of advertising.

References

Anderson, Simon P. and Joshua S. Gans. 2011. "Platform Siphoning: Ad-Avoidance and Media Content." *American Economic Journal: Microeconomics*, 3(4): 1-34.

Chiou, Lesley and Catherine Tucker, 2013. "Paywalls and the demand for news." *Information Economics and Policy*, 25(2), 61-69.

Giorcelli, Michela and Petra Moser, 2016. “Copyright and Creativity: Evidence from Italian Operas.” Working paper.

Goldfarb, Avi, and Catherine Tucker, 2011a. “Privacy Regulation and Online Advertising.” *Management Science*, 57(1), 57-71.

Goldfarb, Avi, and Catherine Tucker, 2011b. “Online Display Advertising: Targeting and Obtrusiveness.” *Marketing Science*, 30(3), 389-404.

Goldfarb, Avi, and Catherine Tucker, 2015. “Standardization and the Effectiveness of Online Advertising.” *Management Science,* 61(11), 2707-2719.

Oberholzer-Gee, Felix and Koleman Strumpf, 2007. "The Effect of File Sharing on Record Sales: An Empirical Analysis." *Journal of Political Economy,* 115(1), 1-42.

Rob, Rafael and Joel Waldfogel, 2006. "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students." *Journal of Law and Economics*, 49(1), 29-62.

Seamans, Robert and Feng Zhu, 2014. Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers." *Management* Science, 60(2), 476-493.

Tucker, Catherine, 2014. “Social Networks, Personalized Advertising, and Privacy Controls.” *Journal of Marketing Research*, 51(5), 546-562.

Waldfogel, Joel, 2012. "Copyright Protection, Technological Change, and the Quality of New Products: Evidence from Recorded Music since Napster." *Journal of Law and Economics*, 55(4), 715-740.

Wilbur, K. C., 2008. “How the Digital Video Recorder Changes Traditional Television Advertising.” *Journal of Advertising*, 37 (1), 143-149.

Table 1: Domain Suffixes

Suffix	**N**	**Location Site Registered (if relevant)**
com	1,503	
uk	158	United Kingdom
net	134	
de	76	Germany
fr	71	France
ro	68	Romania
org	64	
nl	51	Netherlands
hu	42	Hungary
se	32	Sweden
it	29	Italy
ru	27	Russia
be	20	Belgium
dk	20	Denmark
info	19	
ie	17	Ireland
au	16	Australia
tv	16	Tuvalu
br	14	Brazil
pl	11	Poland

Table 2: Countries in Which Traffic Frequently Originates

Country	Count*
United States	2467
United Kingdom	1071
India	750
Canada	577
Germany	418
France	377
Italy	228
Australia	216
Romania	206
Spain	186
Belgium	171
Netherlands	137
Russia	133
Sweden	116
Brazil	110
Pakistan	101
Mexico	87
Austria	79
Japan	79
Algeria	77
Indonesia	74
South Korea	69
Turkey	69
Switzerland	68
China	66

*The count shows the number of times Alexa reports the country among the top 5 places where traffic originates for a website in the data. The top 25 places of origin are shown.

Table 3: Ad Blocking and Rank Change: Various Approaches

	OLS	Censored	IV Regression: Instrument Using:		
			Suffix	Traffic Origin	
				Excl. Missing	Impute Missing
	(1)	(2)	(3)	(4)	(5)
Share of Visitors Using Ad Blocker (P)	0.54706 (0.19363)**	0.63268 (0.20067)**	0.75811 (0.92152)	0.86299 (0.61129)	1.43917 (0.48937)**
N	2,574	2,574	2,574	2,078	2,574

* $p<0.05$; ** $p<0.01$

Notes: The dependent variable is the change in log site rank, June 2013-April 2016, and the independent variable is our measure of the eventual share of site visitors using an ad blocker. Column (1) uses OLS, while column (2) uses a censored approach, since site rank is top-coded at 34 million. Columns (3)-(5) implement IV approaches discussed in the text. Bootstrapped standard error reported in column (5). The bootstrapping procedure repeatedly resamples websites (with replacement). For each drawn sample, we redo the imputation of origin visitor shares for the websites in the sample missing such data, using the average visitor shares at websites with the same suffix in the drawn sample, and then re-run the IV regression. The reported standard errors are the standard deviations of the coefficients from these 1,000 redrawn and re-imputed samples.

Table 4: Ad Blocking and Rank Change: Omitting Websites with Noisy Measurement of Ad Block Rate

	OLS		Censored Regression	
	All Obs.	Omit if LT 10 Visits in PF Data	All Obs.	Omit if LT 10 Visits in PF Data
	(1)	(2)	(3)	(4)
Share Visitors Using Ad Blocker	0.54706 (0.19363)**	0.82065 (0.23227)**	0.63268 (0.20067)**	0.91242 (0.24004)**
N	2,574	2,479	2,574	2,479

* p<0.05; ** p<0.01

Notes: Columns (1) and (3) reproduce results in Table 3. The remaining columns exclude sites where our estimate of the share of site visitors using an ad blocker are based on few observations and therefore potentially measured with error.

Table 5: Site Traffic and Eventual Ad Blocking with the Diffusion of Ad Blocking

	Linear (1)	Linear (2)	Censored (3)	Censored (4)	Linear (5)	Linear (6)
P	-0.2250		-0.2032			
	(0.0990)*		(0.1023)*			
P*(post-1/14)	0.4111		0.4055		0.4111	
	(0.1107)**		(0.1147)**		(0.1145)**	
P*(6/13-1/14)		-0.2250		-0.2034		
		(0.0986)*		(0.1018)*		
P*(1/14-8/14)		0.0593		0.0733		0.2843
		(0.0986)		(0.1030)		(0.1442)*
P*(8/14-3/15)		0.3076		0.3195		0.5326
		(0.0986)**		(0.1031)**		(0.1442)**
P*(3/15-10/15)		0.2024		0.2256		0.4274
		(0.0986)*		(0.1034)*		(0.1442)**
P*(10/15-5/16)		0.1749		0.1956		0.3999
		(0.0986)		(0.1042)		(0.1442)**
Site FE	No	No	No	No	Yes	Yes
N	12,870	12,870	12,718	12,718	12,870	12,870

* p<0.05; ** p<0.01

Note: these estimates relate the change in sites' log ranks to the sites' eventual ad blocking (P). The specifications include five seven-month time periods. The first time period runs from June 2013 through January 2014 and is considered the period prior to widespread ad blocking. The first four columns report linear regressions (cols 1 and 2) and the censored approach explained in the text (cols 3 and 4). The last two columns use the include site fixed effects in linear regressions.

Table 6: Ad Blocking and Rank Change from June 2013
Censored Regressions

	(1)	(2)	(3)
Share Visitors Using Ad Blocker	0.7715 (0.1895)**	0.4865 (0.1947)*	0.6500 (0.1914)**
Alexa Rank in June 2013	-0.0000 (0.0000)**		
Log Alexa Rank in June 2013		-0.1366 (0.0117)**	
June 2013 Rank Deciles:			
2			0.0210 (0.1305)
3			0.0114 (0.1304)
4			0.0080 (0.1306)
5			-0.0657 (0.1306)
6			-0.0911 (0.1306)
7			-0.1160 (0.1309)
8			-0.3772 (0.1309)**
9			-0.5713 (0.1311)**
10 (Worst Rank)			-1.5020 (0.1316)**
N	2,574	2,574	2,574

* $p<0.05$; ** $p<0.01$

Notes: The table reports censored regressions resembling the result in column (2) of Table 3, augmented to include, measures of the site ranks at the start of the sample period (June 2013).

Table 7: Page Views and Rank

Dependent Variable: Log of Websites % of Global Traffic	
Log(Alexa Rank)	-1.0618
	(0.0049)**
Constant	-2.6545
	(0.0455)**
R^2	0.94
N	2,967

* p<0.05; ** p<0.01

Note: Estimated on sample of websites scraped from Alexa's ranking list

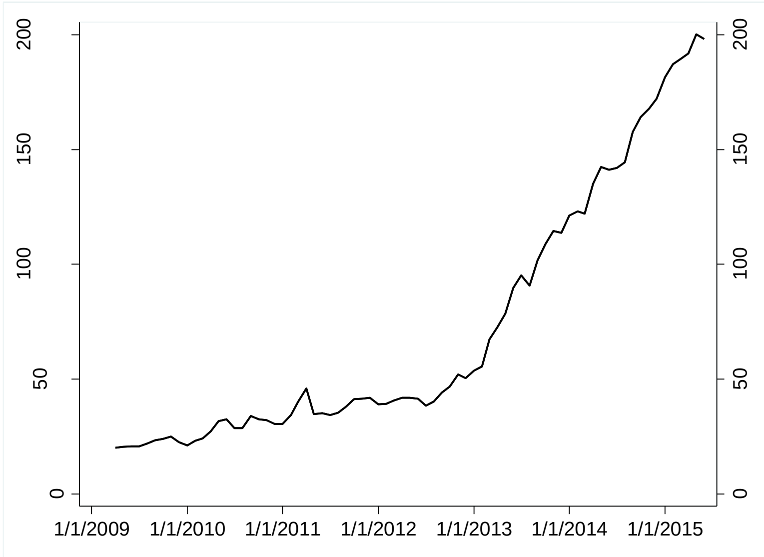

Figure 1
Ad Blocker Global Diffusion

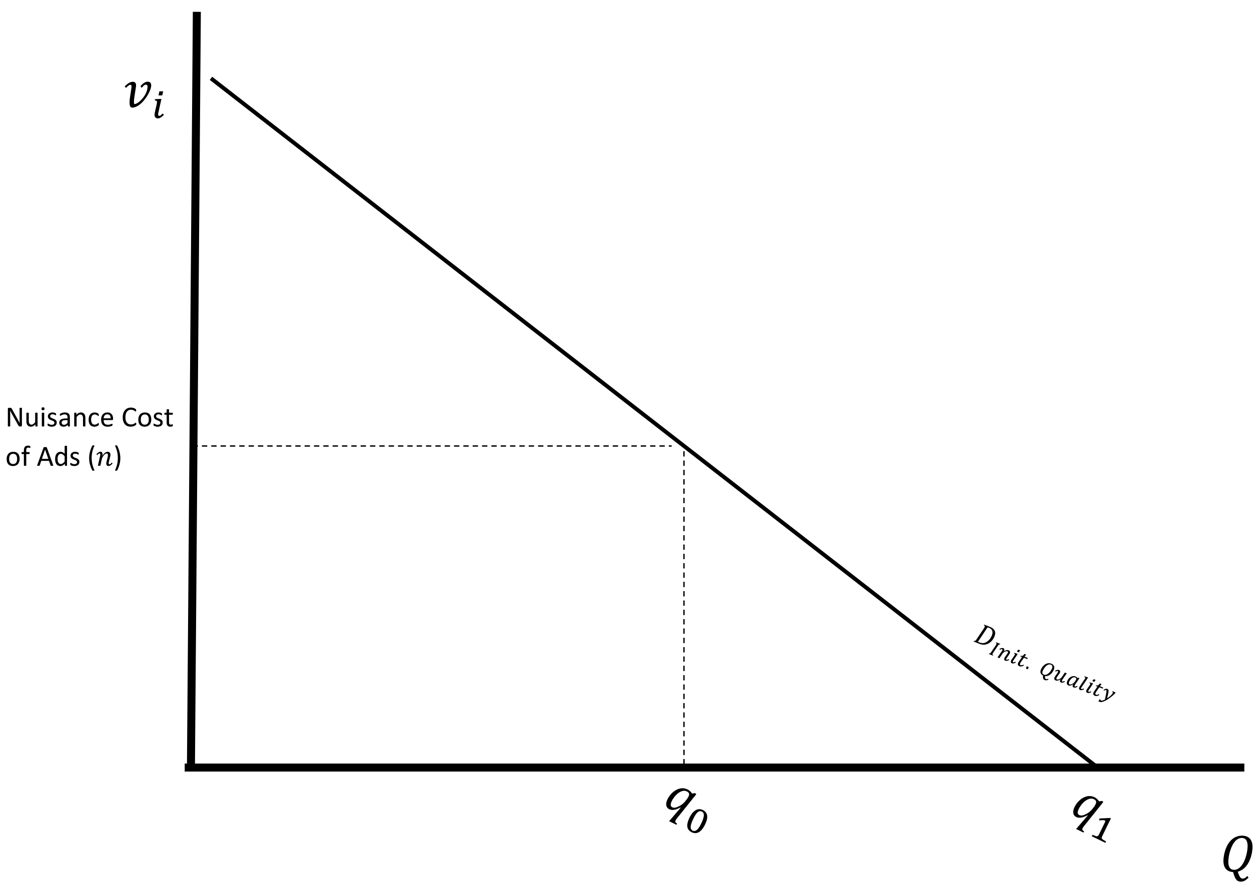

Figure 2
Ad Blocking Graphical Theory: Short Run Impact

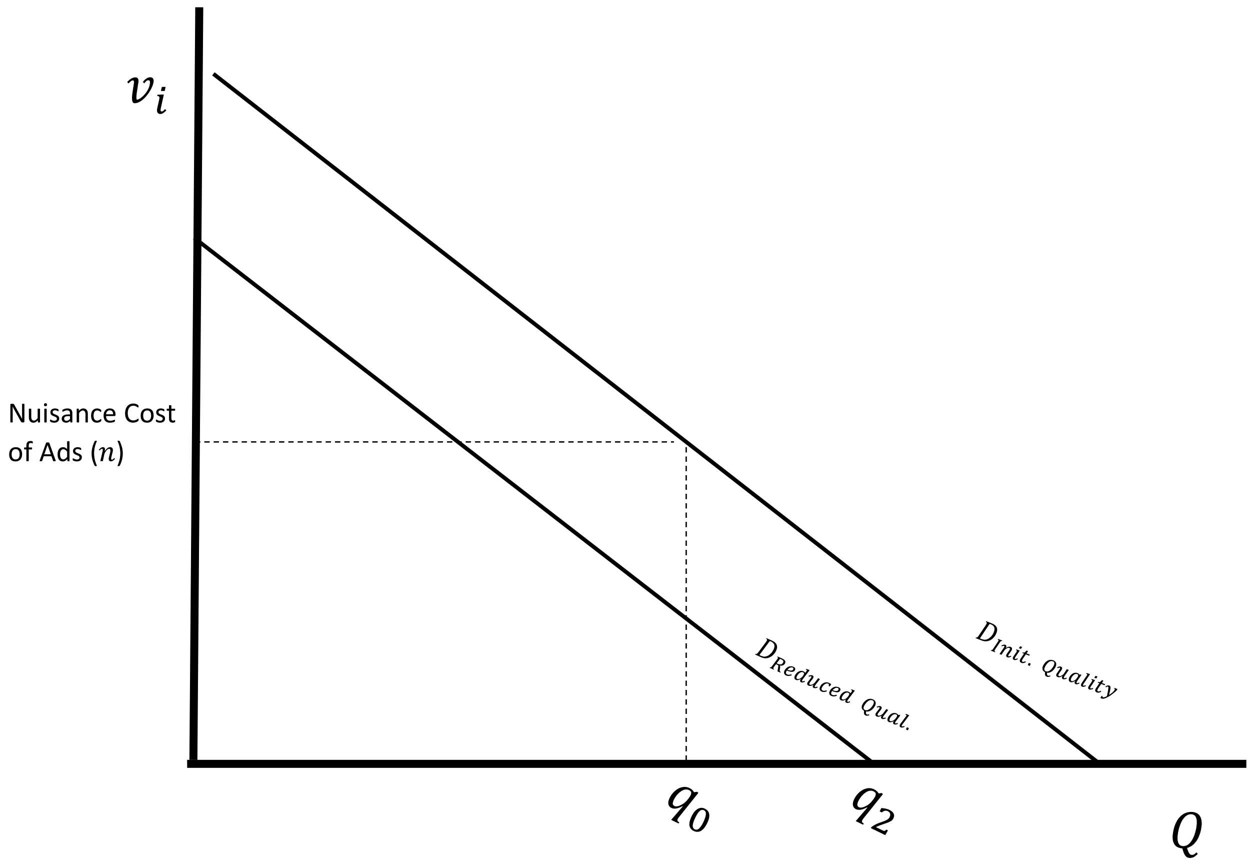

Figure 3
Ad Blocking Graphical Theory: Long Run Impact

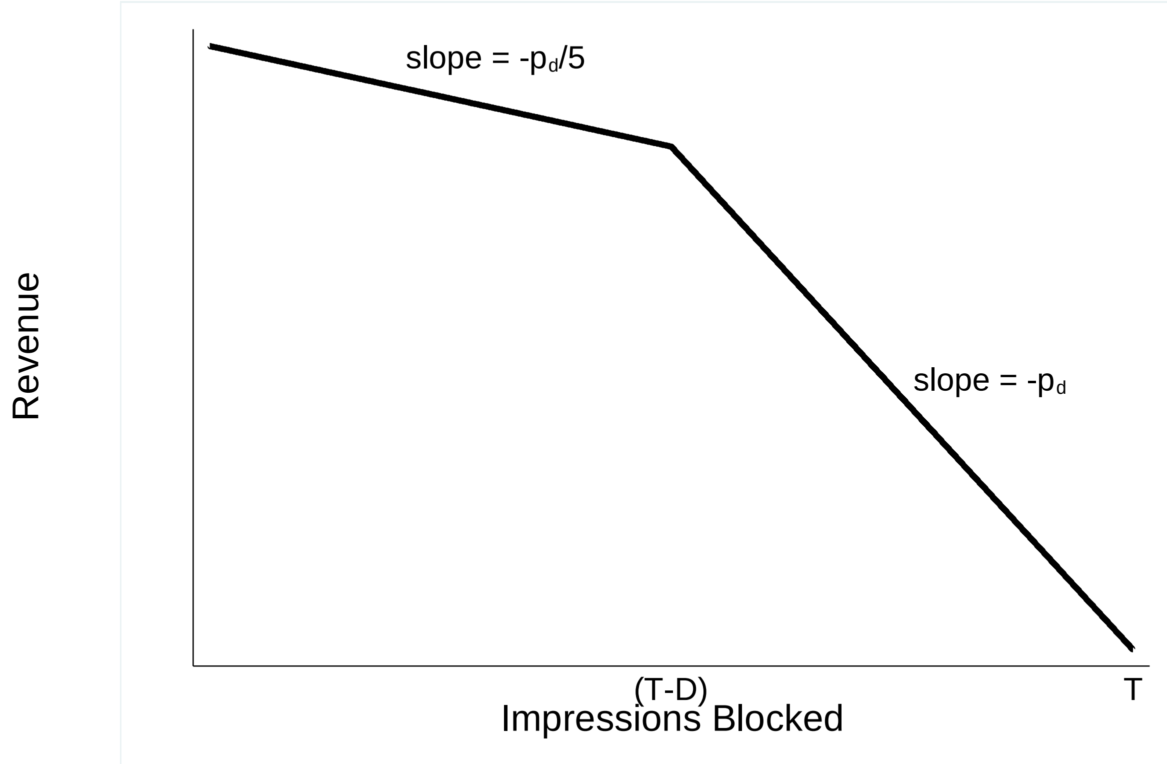

Figure 4
Impact of Blocked Impressions on Revenue

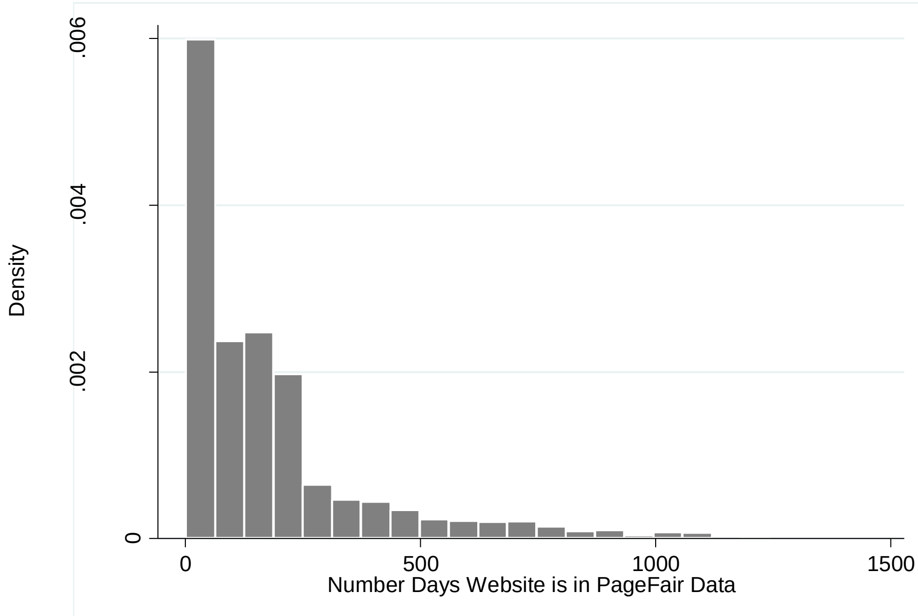

Figure 5
Distribution of Ad Block Rates across Websites

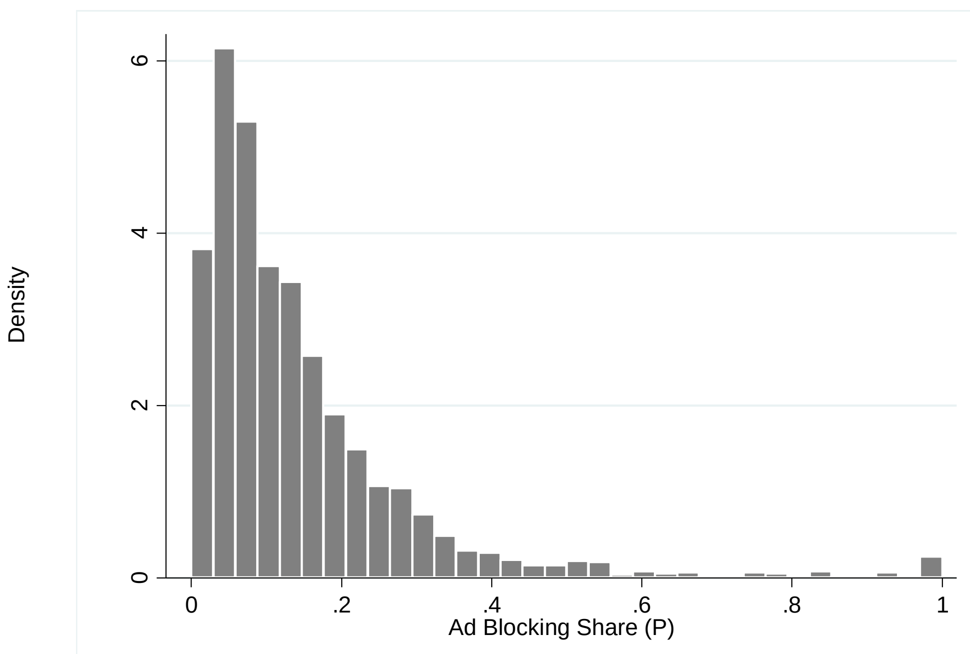

Figure 6
Distribution of Ad Blocking Visitor Share across Websites - P^1 Measure

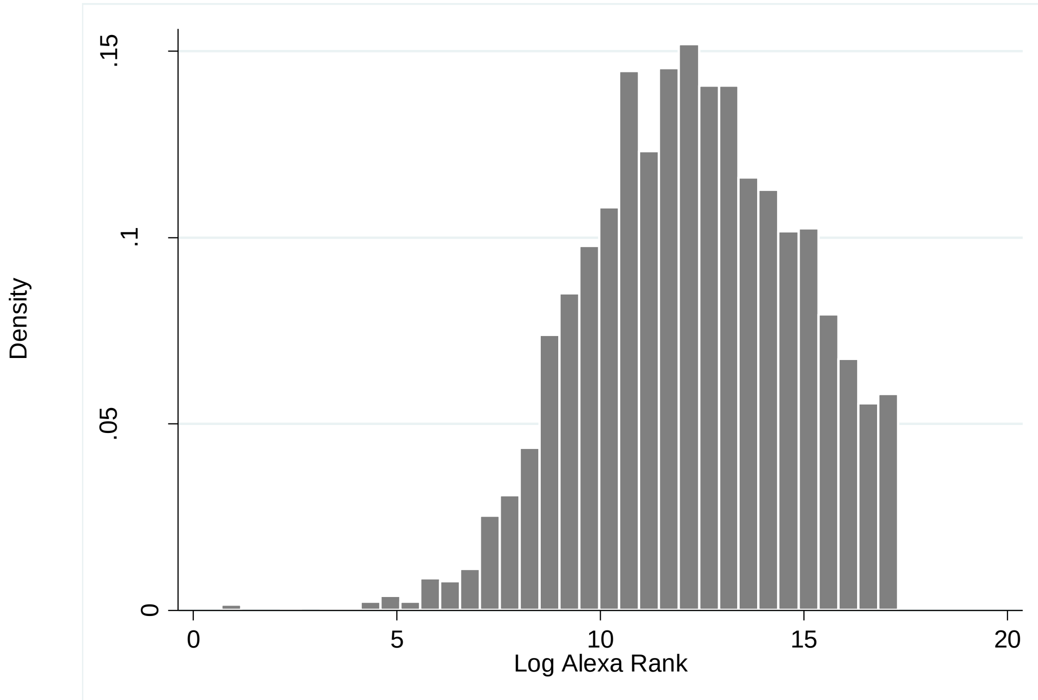

Figure 7
Distribution of Alexa Traffic Ranks in June 2013

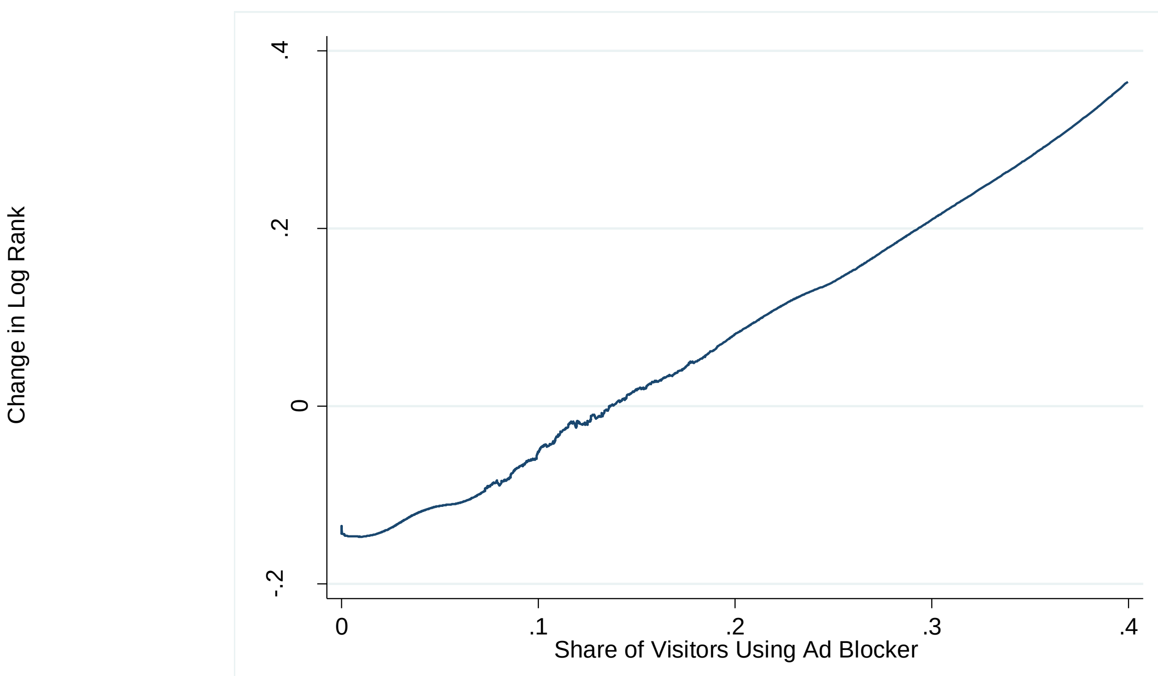

Figure 8
Eventual Ad Block Share and Rank Change

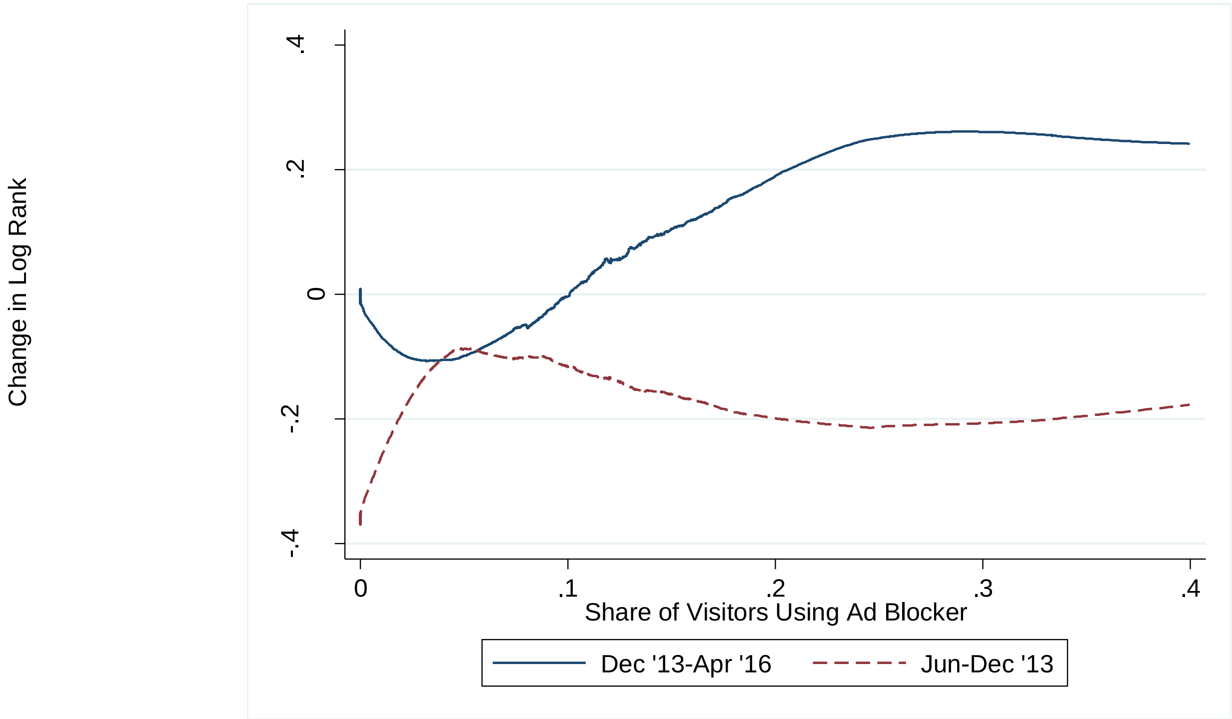

Figure 9
Eventual Ad Block Share and Rank Change
Growth Period vs. Pre-period

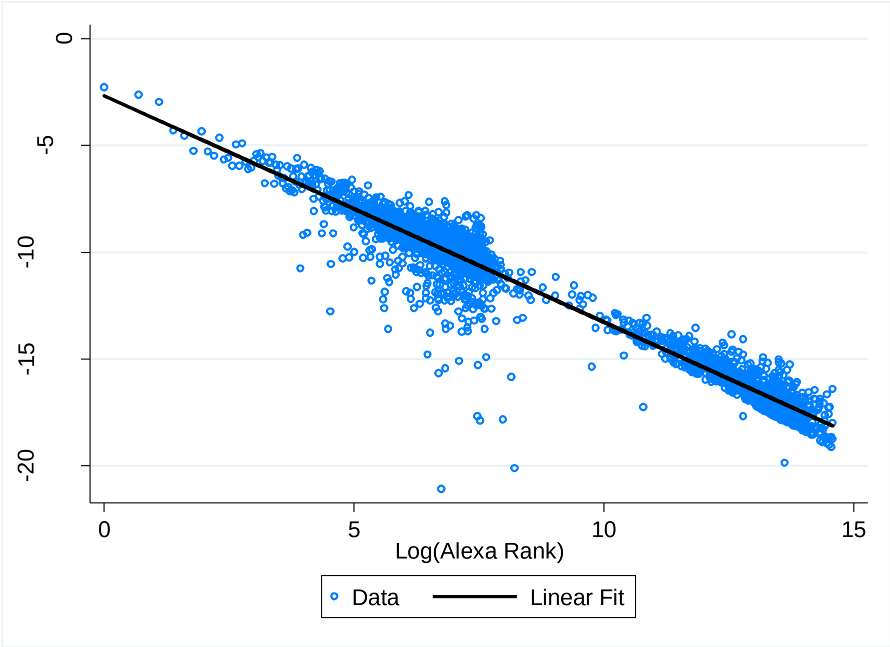

Figure 10
Relationship Between Traffic and Traffic Ranks

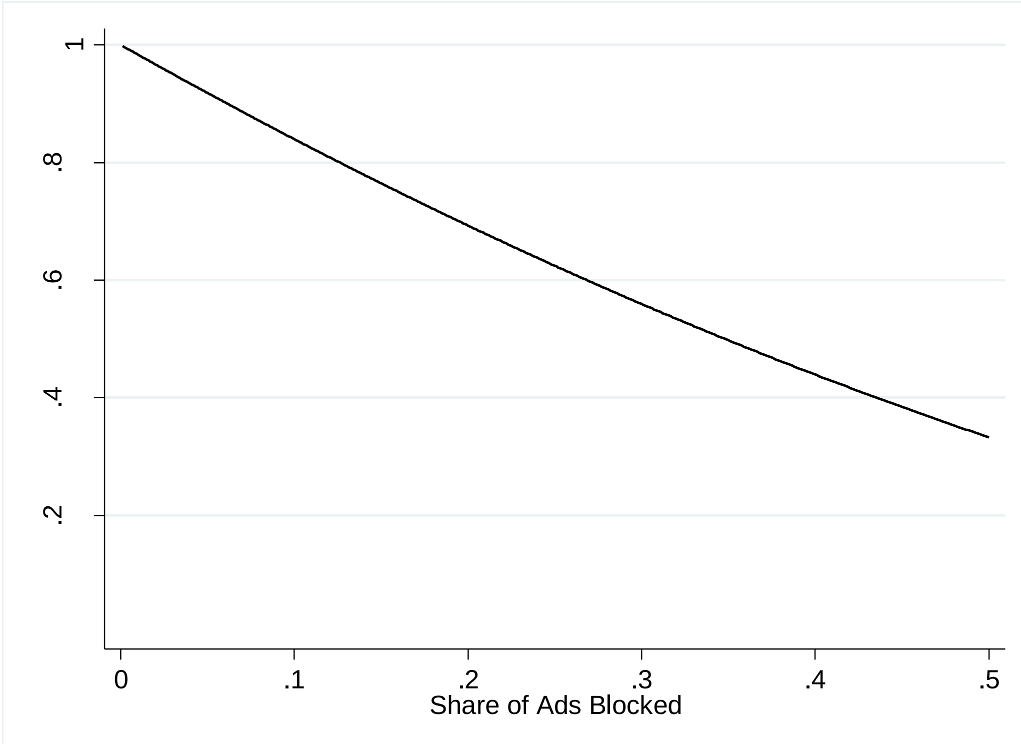

Figure 11
Long Term Impact of Ad Blocking on Revenue

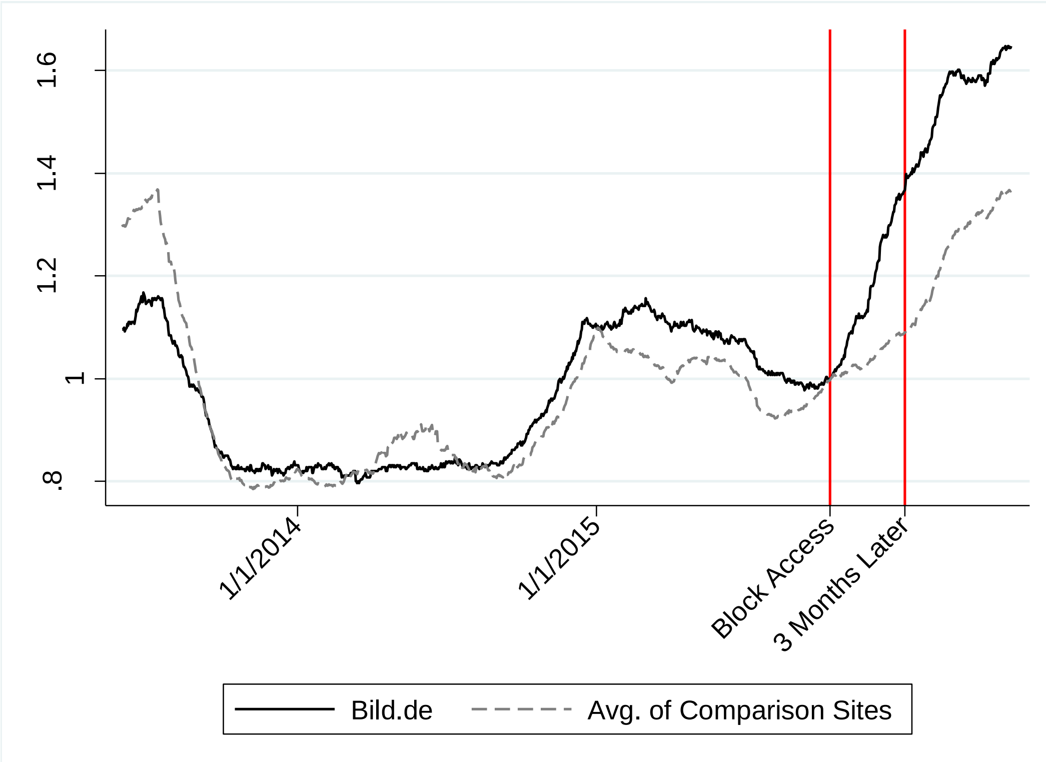

Figure 12
Exiling Ad Blockers and the Impact on Traffic - Case of Bild.de

Appendix – Repeating for Alternative Specification of Share Blocked $\left(P_i^2\right)$

Appendix Table 3

Ad Blocking and Rank Change: Various Approaches

	OLS	Censored	IV Regression: Instrument Using:		
			Suffix	Traffic Origin	
				Excl. Missing	Impute Missing
	(1)	(2)	(3)	(4)	(5)
Share of Visitors Using Ad Blocker	0.64435 (0.23795)**	0.76240 (0.24680)**	0.86816 (1.10606)	1.27566 (0.72553)	1.78642 (0.58503)**
N	2,574	2,574	2,574	2,078	2,574

* p<0.05; ** p<0.01

Notes: The dependent variable is the change in log site rank, June 2013-April 2016, and the independent variable is our measure of the eventual share of site visitors using an ad blocker. Column (1) uses OLS, while column (2) uses a censored approach, since site rank is top-coded at 34 million. Columns (3)-(5) implement IV approaches discussed in the text. Bootstrapped standard error reported in column (5). The bootstrapping procedure repeatedly resamples websites (with replacement). For each drawn sample, we redo the imputation of origin visitor shares for the websites in the sample missing such data, using the average visitor shares at websites with the same suffix in the drawn sample, and then re-run the IV regression. The reported standard errors are the standard deviations of the coefficients from these 1,000 redrawn and re-imputed samples.

Appendix Table 4

Ad Blocking and Rank Change: Omitting Websites with Noisy Measurement of Ad Block Rate

	OLS		Censored Regression	
	All Obs.	Omit if LT 10 Visits in PF Data	All Obs.	Omit if LT 10 Visits in PF Data
	(1)	(2)	(3)	(4)
Share Visitors Using Ad Blocker	0.64435 (0.23795)**	0.82994 (0.25335)**	0.76240 (0.24680)**	0.92820 (0.26178)**
N	2,574	2,479	2,574	2,479

* $p<0.05$; ** $p<0.01$

Notes: Columns (1) and (3) reproduce results in Table 3. The remaining columns exclude sites where our estimate of the share of site visitors using an ad blocker are based on few observations and therefore potentially measured with error.

Appendix Table 5

Site Traffic and Eventual Ad Blocking with the Diffusion of Ad Blocking

	Linear	Linear	Censored	Censored	Linear	Linear
	(1)	(2)	(3)	(4)	(5)	(6)
P	-0.3598		-0.3306			
	(0.1216)**		(0.1257)**			
P*(post-1/14)	0.5916		0.5832		0.5916	
	(0.1360)**		(0.1410)**		(0.1406)**	
P*(6/13-1/14)		-0.3598		-0.3308		
		(0.1211)**		(0.1252)**		
P*(1/14-8/14)		0.0866		0.0995		0.4464
		(0.1211)		(0.1273)		(0.1771)*
P*(8/14-3/15)		0.4060		0.4294		0.7658
		(0.1211)**		(0.1268)**		(0.1771)**
P*(3/15-10/15)		0.2016		0.2337		0.5614
		(0.1211)		(0.1266)		(0.1771)**
P*(10/15-5/16)		0.2329		0.2539		0.5927
		(0.1211)		(0.1282)*		(0.1771)**
Site FE	No	No	No	No	Yes	Yes
N	12,870	12,870	12,718	12,718	12,870	12,870

* $p<0.05$; ** $p<0.01$

Note: these estimates relate the change in sites' log ranks to the sites' eventual ad blocking (P). The specifications include five seven-month time periods. The first time period runs from June 2013 through January 2014 and is considered the period prior to widespread ad blocking. The first four columns report linear regressions (cols 1 and 2) and the censored approach explained in the text (cols 3 and 4). The last two columns use the include site fixed effects in linear regressions.

Appendix Table 6
Ad Blocking and Rank Change from June 2013
Censored Regressions

	(1)	(2)	(3)
Share Visitors Using Ad Blocker	1.0191	0.7036	0.8747
	(0.2331)**	(0.2388)**	(0.2348)**
Alexa Rank in June 2013	-0.0000		
	(0.0000)**		
Log Alexa Rank in June 2013		-0.1378	
		(0.0117)**	
June 2013 Rank Deciles:			
2			0.0055
			(0.1305)
3			0.0049
			(0.1303)
4			0.0002
			(0.1305)
5			-0.0709
			(0.1305)
6			-0.1064
			(0.1304)
7			-0.1312
			(0.1306)
8			-0.3917
			(0.1306)**
9			-0.5891
			(0.1307)**
10 (Worst Rank)			-1.5190
			(0.1315)**
N	2,574	2,574	2,574

* $p<0.05$; ** $p<0.01$

Notes: The table reports censored regressions resembling the result in column (2) of Table 3, augmented to include, measures of the site ranks at the start of the sample period (June 2013).