CESifo WORKING PAPERS

6852
2018

January 2018

Follow The Money: Online Piracy and Self-Regulation in the Advertising Industry

Michail Batikas, Jörg Claussen, Christian Peukert

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 6852*
Category 14: Economics of Digitization

Follow The Money: Online Piracy and Self-Regulation in the Advertising Industry

Abstract

In this paper, we study the effects of a self-regulatory effort, orchestrated by the European Commission, that aims to reduce advertising revenues for publishers of copyright infringing content. Historical data lets us follow how the third-party advertising and tracking services associated with a large number of piracy websites and a corresponding set of legitimate "placebo" websites change after the agreement to self-regulate went in place. We find that larger EU-based advertisers comply with the initiative and reduce their connections with piracy websites. We do not find reductions for other non-advertising services that track consumers, which has potentially important implications for the efficiency of targeted advertising.

JEL-Codes: K400, L500, L800.

Keywords: piracy, copyright enforcement, online advertising, natural experiment.

Michail Batikas
LMU Munich
Munich / Germany
m.batikas@lmu.de

Jörg Claussen
LMU Munich
Munich / Germany
j.claussen@lmu.de

*Christian Peukert**
UCP – Católica Lisbon School of Business and Economics
Lisboa / Portugal
christian.peukert@ucp.pt

*corresponding author

January 9, 2018
The paper has benefited from feedback of conference participants at Workshop on Economics, Law and Policy of Communications and Media Florence, ITS Europe Conference Passau, TILEC-GovReg Workshop on Economic Governance of Data-driven Markets Tilburg, and seminars at LMU Munich, Northeastern and DIW Berlin. We acknowledge the support from FCT - Portuguese Foundation of Science and Technology for the project UID/GES/00407/2013.

1 Introduction

Advertising is an important institution in multi-sided markets. On the one hand, it enables firms to inform and persuade consumers about their products. On the other hand, it generates incentives for the creation of content. The classic example is the newspaper market, where advertisers fund journalism by effectively subsidizing the price charged to readers. It is well understood that in this market, as well as in many other media markets, the interests of advertisers and consumers are not necessarily aligned, which potentially affects the type of content created by publishers (see e.g. Xiang and Sarvary, 2007; Ellman and Germano, 2009).

However, less well understood are the potentially important externalities that make ad-funded content dynamically welfare decreasing, although it creates a positive surplus for all directly involved parties (advertisers, consumers and publishers) in the short run. To illustrate this idea, consider the recent phenomenon of false and misleading information ("fake news"), which – anecdotally – is mainly produced to create an audience for advertisers.[1] The observation that a number of news articles about US politics, which are based on objectively false information, where widely shared on social media created a public debate on whether fake news could have biased voting decisions in the 2016 presidential election.[2] Just a few days after the election, major players in the advertising industry, including *Google*, *Facebook* and *AppNexus* announced that they will introduce mechanisms to ban publishers of fake news and hate speech from using their services.[3] This seems to be in line with a literature that has argued that firms can have incentives to strategically internalize ("self-regulate") to preempt political action (e.g. Maxwell et al., 2000), or sacrifice profits as a response to consumer activism (e.g. Hendel et al., 2017).[4] However, anecdotally it seems that the strategic incentives to self-regulate can vary across firms. Considerably

[1] See http://www.bbc.com/news/magazine-38168281.

[2] Evidence in Allcott and Gentzkow (2017) suggests that the average US adult was exposed to about one fake news article during the election period. The authors argue that this is too little exposure to change the election outcome, i.e. relative to the margin with which Donald Trump won.

[3] See for example http://www.wsj.com/articles/google-pulled-into-debate-over-fake-news-on-the-web-1479159867.

[4] Political action is not unlikely in this market. A recent example is the German *Netzwerkdurchsetzungsgesetz*. Coming into effect in October 2017, the legislation forces platforms with more than two million users to remove content containing hate speech or other criminal material within 24 hours. See http://www.bbc.com/news/technology-40444354.

smaller firms such as *Pubmatic* and *OpenX* announced that they will not change their policies, arguing that "we do not make it our business to police editorial content" and "[we are] proud to support a free and vibrant internet".[5] Self-regulation is not uncommon in the advertising industry and its underlying economics appear to be more general and independent of the context of elections.[6]

In this paper, we study the strategic choice of whether or not firms in the online advertising industry follow a self-regulation effort in the context of copyright infringing content ("online piracy"). We specifically investigate heterogeneity in terms of firm size and firm type. According to a joint report of the music and advertising industry published in 2012, more than two thirds of copyright infringing websites are predominately financed by advertising.[7] When advertising makes it profitable to run a platform that makes unlicensed software and media products available to consumers, and consumers substitute away from licensed products, the negative externalities of advertising appear in the form of reduced revenues of creators. In the long run, this can reduce incentives to invest in the creation of such products (e.g. Bae and Choi, 2006).

We identify the strategic reactions to self-regulation in a setting where a natural experiment, based on geography, increases the threat of public regulation for some firms in the advertising industry, and not to others. We exploit the orchestration of self-regulation by the European Commission, in which EU-based advertising firms face pressure to establish a voluntary agreement to reduce the flow of revenues to IP-infringing websites. The EC explicitly states that "[...] those codes of conduct at EU level could be backed by legislation, if required, to ensure their full effectiveness."[8] Specifically, we compare the involvement of EU-based and non-EU-based firms in the piracy market, before and after an agreement on basic principles of self-regulation was signed on October 21, 2016.

We collect unique data that allows us to observe the ad and tracking services associated

[5]See http://adage.com/article/digital/major-ad-technology-company-bars-breitbart-news-hate-speech/306887/.

[6]For example, there are several industry associations dealing with self-regulation, like the *Advertising Self-Regulatory Council* in the US, the *European Advertising Standards Alliance* and the *Advertising Standards Bureau* in Australia. Examples of self-regulation include advertising to children, food and beverage advertising, and gambling advertising; see https://adstandards.com.au/products-issues.

[7]Google and PRS (2012), "The six business models for copyright infringement", https://docs.google.com/file/d/0Bw8Krj_Q8UaENDhEOG1LVFRhVkU/view.

[8]See https://ec.europa.eu/growth/industry/intellectual-property/enforcement_en.

with 269 piracy websites at various points in time, based on the HTTP requests that these websites make to third parties. For each piracy website, we find two legitimate ("placebo") websites that are similar in audience size, and collect information on the associated third parties. This allows us to follow 923 third-party services and their interactions with piracy and placebo websites from early 2016 to early 2017. Finally, we add meta information that lets us distinguish between EU-based and non-EU-based, smaller and larger, and ad and non-ad ("tracking") services. We estimate parameters of reduced form difference-in-differences models, allowing for group-specific time trends and individual-level fixed effects.

Two main results emerge from our analysis. First, our results suggest that EU-based ad services on average do not seem to comply with the self-regulation. We find a small and statistically not significant change in the number of piracy websites per ad service. Once we allow for heterogeneity in size, however, we find that the number of piracy websites served by an ad service is significantly decreasing in the size of the ad service. We interpret this as evidence that the likelihood of complying with the self-regulation increases with firm size. We further provide some evidence that ad services that were not active in the piracy market before the self-regulation agreement increase their presence on piracy websites afterwards. Second, we contrast ad services and other non-directly ad-related services ("tracking services"). We find that the number of piracy websites associated with tracking services does not change significantly. We interpret this as evidence that it is mainly ad services that comply to the self-regulation.

These results raise concerns about the overall effectiveness of the self-regulation effort with respect to reducing incentives for publishers to supply unlicensed content. Somewhat consistent with this view, we show that there is at least no clear downward trend in the rate of entry of new piracy websites after the agreement, and in the number of visits to existing piracy websites.

We believe that our results allow to draw at least two broader implications. First, our results show that creating incentives for firms to self-regulate and internalize an externality can be intricate when firms are heterogenous. Second, we argue that the fact that tracking services don't seem to respond to incentives to take part in a self-regulation can increase the efficiency of targeted advertising. The information that a consumer has visited a piracy website can be valuable to advertisers seeking to advertise elsewhere. For example, knowing that a consumer has navigated to an unlicensed live sports streaming website can be indicative of that consumer's interest in sports. In other cases, when visiting a piracy website signals that a consumer has a low willingness to pay, data on the browsing history can be valuable for negative targeting, i.e. to make sure that this user is not served the ad. Because of these informational externalities, incentivizing self-regulation, even when relatively ineffective with respect to its main goal, can lead to more efficient market outcomes than an one-size-fits-all regulation that affects the entire industry.

Our study makes contributions to several literatures. First, we provide evidence that different types of firms face different types of trade-offs in the strategic decision to self-regulate, adding to a large literature on corporate self-regulation (see e.g. Maxwell et al., 2000; Egorov and Harstad, 2017). Second, we contribute to the literature on the economics of online advertising, tracking and privacy (see Tucker, 2012; Goldfarb, 2014; Peitz and Reisinger, 2015 for summaries). Third, we add to the growing literature that evaluates IP enforcement policies (Adermon and Liang, 2014; Danaher and Smith, 2014; Reimers, 2016; Peukert et al., 2017; Aguiar et al., 2017).

2 Context

2.1 The market for online advertising and consumer tracking

The market for online advertising and consumer tracking is comprised of the system that connects advertisers to consumers via publishers. While publishers can in principle directly connect to advertisers, in practice, a large majority of publishers uses intermediaries to do so. Over the last 20 years, the online advertising industry has grown by a factor of 50 in terms of number of distinct services on the market, and by a factor of 4 in terms of the median number of distinct services used per website (Lerner et al., 2016). Now, about 88% of the top 1 million most popular websites operate with at least one third-party service, and out of those, the average site requests about 9 distinct third-party domains (Libert, 2015). At the same time, the industry has become considerably concentrated with the top

20 services covering about a third of the market (Gill et al., 2013; Schelter and Kunegis, 2016). *Advertising networks*, such as Google's DoubleClick, take a central position in this system in that they connect a range of auxiliary advertising services that intermediate and facilitate a market between advertisers, publishers and data providers. Advertisers and publishers sometimes buy services (*demand-side* and *supply-side platforms*) that enable more effective targeting, most prominently from *tracking services* which collect and consolidate behavioral data of users. For example, a tracking-plugin installed on a website places a cookie on a consumer's machine, which allows the publisher to follow the same consumer within her platform (e.g. Google Analytics), but also allows the third-party to track consumers across other websites that operate the same plugin. Prominent examples of across-publisher tracking services are the "like" and "share" buttons operated by social media websites (Chaabane et al., 2012; Roosendaal, 2012). These data are sold to advertisers or advertising networks, which use it to predict the "value" of a consumer, and which consequently affects their willingness to pay for the consumer's eyeballs.[9] It is important to note that firms are often vertically integrated and assume multiple roles at the same time. For example, an advertising network can collect data from the tracking of consumers, use this data to enhance its prediction models, but also sell it to advertisers or other advertising networks.

Much of what is going on in real-time as a user visits a publisher's website is not directly visible for the consumer, nor are the identities of the involved parties that collect and trade personal information completely transparent. As a result, a growing body of literature in economics and related fields has focused on privacy issues (see Acquisti et al., 2016 for a summary).

2.2 The economics of online advertising and consumer tracking

The literature on the economics of online advertising and consumer tracking has been steadily growing (see Tucker, 2012, Goldfarb, 2014 and Peitz and Reisinger, 2015 for

[9] For example, Facebook states in a 2015 press release: "Last year, we introduced online interest-based advertising – ads based on people's use of other websites and apps [...]. For example, with online interest-based ads, if you visit hotel and airline websites to research an upcoming trip, you might then see ads for travel deals on Facebook.", see https://www.facebook.com/notes/facebook-and-privacy/a-new-way-to-control-the-ads-you-see-on-facebook/926372204079329/.

summaries). In what follows we review a selection of studies that are especially useful given the empirical context of our study.

The model of de Cornière and de Nijs (2016) provides some intuition why a publisher would allow advertisers to get access to detailed information about her users. The authors look at the problem of a monopoly publisher that can decide on whether to install a tracking service. In their model, the tracking service enables advertisers to charge higher prices on the product market. However, only with sufficient competition across advertisers, the publisher can fully capture this rent, which drives incentives to install the tracking service.

A second set of papers helps to see under which circumstances competing publishers would share information about their users by using the same or overlapping third-party services. In the model of Athey et al. (forthcoming), there are two competing publishers that sell ad space to heterogeneous advertisers. Both publishers can perfectly track consumers on their own platform, but not beyond. With multi-homing consumers, advertising is inefficient in that multi-homing advertisers may not be able reach some consumers and may advertise too often to others. As a result, low-type (low valuation per consumer) advertisers single home, high-type advertisers multi home (because of opportunity cost of not informing consumers). In the set-up of Ghosh et al. (2015), publishers can decide to share their information about consumers with other publishers. In this model, when advertisers value consumers homogeneously, all publishers agree, i.e. either all will decide to share information, or none. However, with heterogeneous advertisers, in the sense that some advertisers value high-type consumers more than other advertisers, sharing information makes some publishers better off, some worse off. This is because advertisers can better distinguish between high-type and low-type consumers visiting the publisher, which increases their average willingness to pay, reducing demand for ads at competitive publishers. Put differently, because of the leakage effect of information sharing, the advertiser will be able to target the same consumer elsewhere at a lower price. In the model of D'Annunzio and Russo (2017), publishers implicitly share information about consumers by joining an advertising network. In their model, consumers can decide whether they want to share their information with publishers. The authors show that the possibility to opt out of tracking (e.g. by deleting cookies), can be welfare decreasing, because consumers ignore the positive externality of tracking on advertising revenues.

Empirically, there seems to be important heterogeneity with regards to how valuable the information from consumer tracking is to advertisers. Lambrecht and Tucker (2013) show that the effectiveness of targeted advertising, using information on the previous browsing history of a consumer ("retargeting"), largely depends on the types of websites that the consumer has visited before. Relatedly, Tucker (2014) and Bleier and Eisenbeiss (2015) show that the effectiveness of ads depends on how much information (e.g. about browsing history) the advertiser uses to personalize the ad.

2.3 Evidence on the effectiveness of online copyright enforcement

In the last decade, firms and governmental authorities have used a variety of different approaches to enforce copyright on the Internet. The aim has been to counter the potentially large negative revenue effects of online piracy, which may reduce long-run incentives to invest in the creation of content. A growing body of literature evaluates these policies and arrives at sobering conclusions.

A first set of actions was targeted at the demand-side, involving stricter anti-piracy legislation, and blocking access to certain websites. According to McKenzie (2017), the announcement and introduction of stricter laws in France, New Zealand, South Korea, Taiwan, UK, and the US did not affect movie revenues at the box office. For digital music sales, however, Danaher et al. (2014) report an increase of 22-25% as a result of a new law in France. Adermon and Liang (2014) find a similar effect caused by stricter legislation in Sweden, but they also show that the effect diminishes after 6 months. Evaluating the effects of DNS-blocking in the Netherlands, Poort et al. (2014) find a small decrease in self-reported piracy-levels of music, video, games and books. But also in their data, the effect vanishes after 6 months for most content types.

A second line of research investigates the effects of supply-side measures, such as the takedown of specific content, entire hosting services and link directories. Reimers (2016) finds that private efforts to remove infringing content increases sales of electronic books by 14%. Danaher and Smith (2014) find that the number of downloads of licensed movie content increases by 7-9% after the shutdown of Megaupload. In the same context, Peukert et al. (2017) find a similarly sized effect on the theatrical revenues of widely released blockbuster movies, however, revenues of the average movie even decline. The authors argue that piracy can have positive externalities through word-of-mouth, which are more prevalent for movies with smaller advertising budgets. Evaluating the shutdown of a link directory, Aguiar et al. (2017) use individual-level clickstream data to show that a reduction in consumption of unlicensed content is relatively small, mainly because consumers quickly switch to other piracy websites. Further, they don't find much evidence for an increase in the demand for licensed services.

As a result, it is not surprising that online piracy is still prevalent despite the plethora of enforcement efforts: According to a report published by the UK Intellectual Property Office, a consumer survey from the first half of 2016 suggests that of all consumers who consumed online entertainment content in past three months, 25% have done so at least once via unlicensed channels.[10]

2.4 "Follow the money" and self-regulation in online advertising

Perhaps by recognizing the weaknesses of demand-side and supply-side efforts, a new approach to online copyright enforcement has gained popularity in recent years. The idea is that self-regulation of the advertising industry, in which intermediaries in the online advertising ecosystem commit to not serving ads or ad-related services to websites that contain IP infringing content. This approach is largely born out of the observation that the economic incentives to operate a website that hosts copyright infringing content are to a large extent driven by advertising revenues. According to industry reports, more than two thirds of copyright infringing websites are predominately financed by advertising,[11] and annual ad revenues of piracy sites are in the range of $230 million, with profit margins of 80%.[12] Commonly referred to as "follow the money", initiatives have been started in a variety of countries since 2014.[13] In the United Kingdom, a private-public

[10] Intellectual Property Office (2016), "Online Copyright Infringement Tracker, Latest wave of research Mar 16–May 16, Overview and key findings", p. 19.

[11] Google and PRS (2012), "The six business models for copyright infringement", https://docs.google.com/file/d/0Bw8Krj_Q8UaENDhEOG1LVFRhVkU/view.

[12] See "Good money gone bad – Digital thieves and the hijacking of the online ad business", available at https://drive.google.com/file/d/1W7FB6deYUzqMzJe3TBuNgtpUIlWrLyP1.

[13] See the chronology detailed in "Copyright enforcement online: policies and mechanisms", issued by the industry organization European Audiovisual Observatory, 2015, available at https://drive.google.com/

consortium of copyright owners and the City of London Police's Intellectual Property Crime Unit has started to provide a list of websites for advertisers to avoid.[14] In Italy, France and the US, representatives of the advertising industry have announced that they want to introduce self-regulatory efforts, with (non-legislative) support of the respective governments.[15] These approaches were taken to a supranational-level when the European Commission announced to "facilitate the development of further voluntary Memoranda of Understanding to reduce profits of commercial scale IP infringement in the online environment" in a communication from July 2014.[16] The intention was further underlined in a communication from December 2015, where the Commission stated to "take immediate action to engage, with all parties concerned, in setting up and applying 'follow-the-money' mechanisms, based on a self-regulatory approach", and with the important addition that "codes of conduct at EU level could be backed by legislation, if required to ensure their full effectiveness."[17] On March 14, 2016, a stakeholder meeting was held to discuss "the possibility of establishing a voluntary agreement at EU level in order to avoid the misplacement of advertising on IP-infringing websites, thereby restricting the flow of revenue to such sites while safeguarding the reputation of the advertisers and the integrity of the advertising industry."[18] According to the EC's communication, this meeting brought "all the interested parties together (the advertising industry, intermediaries, content protection sector, online media, right owners, civil society, consumer organisations, brands and advertisers)." Finally, on October 21, 2016, stakeholders agreed on the basic principles of an agreement. The "Guiding Principles" explicitly name the purpose of dissuading the placement of advertising on IP infringing websites and apps, thereby minimizing the funding of IP infringement through advertising revenue. The text further states that the European Commission will set up an institution that will allow for the objective measuring

file/d/1lj18CGG6tSzDpBoOLCv3SVwA-uC7f7Hq.

[14] See https://www.theinquirer.net/inquirer/news/2352052/uk-ip-police-promise-to-disrupt-advertising-on-illegal-websites.

[15] See for example http://www.ifpi.org/news/00-FIMI-announcement and https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/IPEC/2016jointstrategicplan.pdf.

[16] See "Towards a renewed consensus on the enforcement of Intellectual Property Rights: An EU Action Plan", p. 6, available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014DC0392.

[17] See http://ec.europa.eu/newsroom/dae/document.cfm?action=display&doc_id=12526.

[18] See https://ec.europa.eu/growth/industry/intellectual-property/enforcement_en

and reporting of the effectiveness of the agreement.[19]

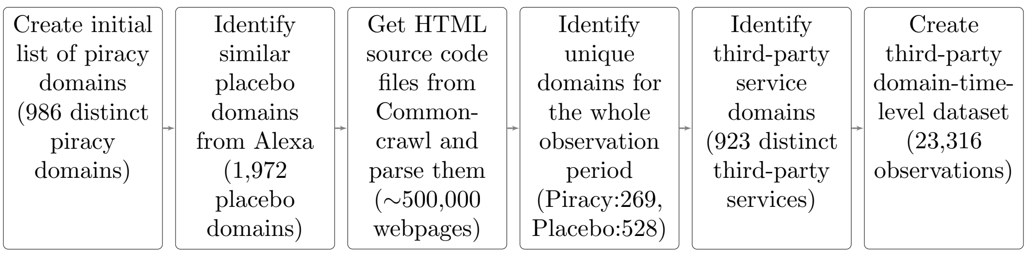
Our empirical analysis below is based on the idea that this form of public-private self-regulation, that comes with a substantial threat of public-only regulation, is exogenous from the perspective of the individual firm. It seems especially unlikely that an individual firm, nor the European advertising industry as a whole could have affected the exact timing of the agreement on October 21, 2016. We will provide a wide range of econometric tests and counterfactual exercises to challenge the exogeneity assumption.

3 Data and identification

In what follows, we first describe the multi-stage data collection process and then illustrate the data using some descriptive statistics. Finally, we introduce the identification strategy and model specification.

3.1 Data collection

Figure 1: Data collection procedure

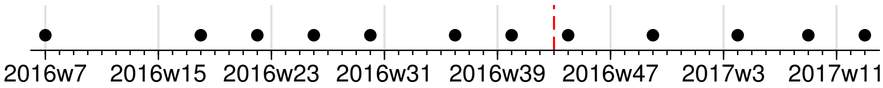


The data collection procedure is summarized in Figure 1. As a first step, we create a list of domains that are associated with unlicensed content, such as the ability to download and stream music, films, live TV, and software. Using a variety of sources, including regularly updated blacklists provided by IT security firms, such as *Shalla*, and lists of websites in industry reports,[20] we arrive at 986 distinct piracy domains. We then construct a sample of legitimate websites that reach a similarly sized audience. Using the ranking of the top 1 million most frequently visited websites provided by market research firm *Alexa*, for each

[19] The text of the Guiding Principles can be found here: https://ec.europa.eu/docsroom/documents/19462/attachments/1/translations/en/renditions/native.

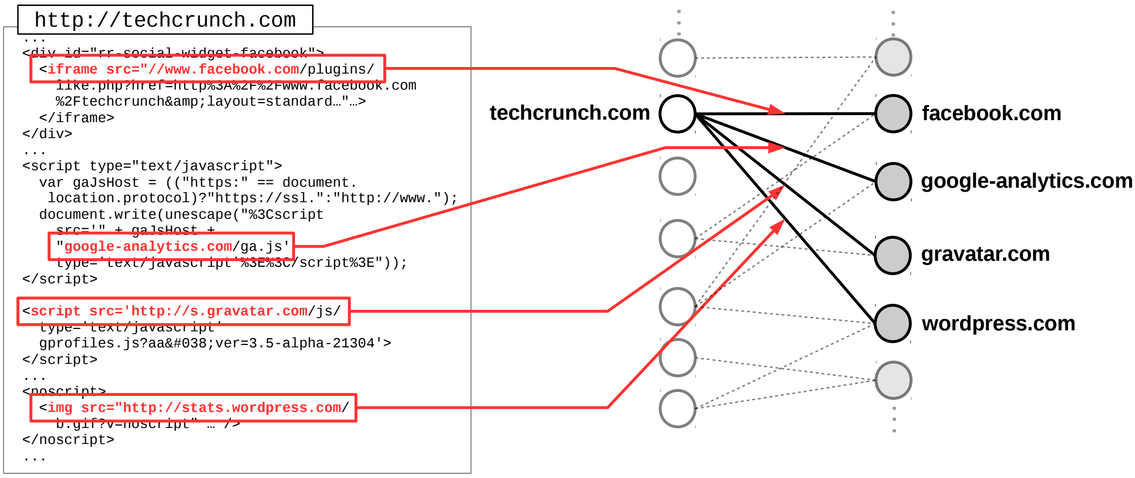
[20] For example we use a list in the appendix of "Good money gone bad – Digital thieves and the hijacking of the online ad business", available at https://drive.google.com/file/d/1W7FB6deYUzqMzJe3TBuNgtpUIlWrLyP1.

Figure 2: Timing of data snapshots



Note: *Available data snapshots from Common Crawl. The vertical line indicates the week of October 21, 2016.*

Figure 3: Example of third-party requests



Note: *Similar to Figure 1 in Schelter and Kunegis (2016).*

piracy domain, we find 1,972 "placebo"domains that are placed one position above or one position below the piracy domains from our initial list of piracy domains. Piracy and placebo domains together define our gross sample of websites for which we want to track the associated third-parties over time. We obtain archived versions of the HTML source code of all URLs for each domain in our gross sample from *Common Crawl*, a project that has crawled billions of webpages periodically since summer 2013. Notice that we obtain the HTML source code of all subdomains for each domain and not only the HTML source code of the landing page. Not all domains in our gross sample are consistently indexed in the Common Crawl corpus, such that we limit our sample to those domains that we observe throughout our whole observation period. This reduces the number of distinct domains to 807, out of which 269 are piracy domains. As is depicted in Figure 2, our first observation week is week 7 of 2016, the last observation week is week 13 of 2017. The frequency of data snapshots is not perfectly balanced – on average we observe a datapoint every 5.3 weeks. The vertical line in Figure 2 indicates the week of the "experiment", showing that we observe 7 snapshots before and 5 snapshots after.

We then access the historical versions of the HTML pages from a local instance of a web server with PhantomJS, a headless browser emulation that allows to monitor all requests to third-party services that a webpage makes, including those after executing client-side scripts.[21] Figure 3 illustrates a stylized example of third-party requests that are hard-coded in the HTML code.

To be able to distinguish between advertising-related and non-advertising-related third party requests, we match the domains of third-party requests to a list of known advertising services provided by market research firm *Evidon.* Around 20% of the third-party domains in our sample are pure advertising services, i.e. the ones that "provide advertising or advertising-related services such as data collection, behavioral analysis or retargeting" or "deliver advertisements that generally appear on adult content site".[22] The three tracking domains in the advertising category most used by piracy domains in our sample are: `doubleclick.com` (Google), `adnxs.com` (AppNexus) and `bluekai.com` (Oracle).

We infer the geographical location of a third-party service from its top level domain (TLD). That is, we assume that a domain that ends with *.co.uk* is operated by a firm legally based in the United Kingdom, *.de* in Germany, *.fr* in France, etc.[23]

For each of the 3,043 third-party domain in our sample, we construct a time-invariant measure of its overall popularity using the extensive data set collected by Libert (2015). Libert studies the third-party services connected to 800,863 websites from a consumer protection perspective. Our measure of popularity is the share of websites that are connected to a specific third-party domain relative to the total number of websites in Libert's data.

Finally, we aggregate our dataset to the third-party-domain-snapshot-level, by counting on how many piracy domains (on how many placebo domains) we observe a given third-party domain at least once in a given time snapshot.[24]

Table 1: Descriptive statistics on third-party domains

	Non-EU (N=21,347)		EU (N=1,969)		Total (N=23,316)	
	Mean	SD	Mean	SD	Mean	SD
Number Piracy Websites	2.346	12.961	0.790	1.258	2.215	12.416
Number Placebo Websites	3.529	22.398	1.187	6.654	3.332	21.530
Popularity (%)	0.478	3.215	0.032	0.140	0.440	3.079
After	0.439	0.496	0.454	0.498	0.440	0.496
Not Served	0.467	0.499	0.416	0.493	0.463	0.499
Ad-Service	0.214	0.410	0.113	0.317	0.205	0.404

3.2 Descriptive statistics

The average third-party domain in our sample is connected to 2.2 piracy websites (0.81% of all piracy sites in our data), and connected to 3.3 placebo websites (0.61% of all placebo sites in our data). The market seems largely concentrated. The third-party domain with the largest number of connected piracy sites (267 or 99.2%), and with the largest number of connected placebo sites (528 or 100%) is `google.com`. Looking at overall popularity, we can draw similar conclusions. The most popular third-party domain, `google-analytics.com` is connected to 69% of the 800k websites in the Libert (2015) data. The median third-party is connected to 0.007% of the websites, the average third-party domain is connected to 0.44% of the websites. About 44% of our observations are after October 21, 2016 (see also Figure 2). As indicated by the dummy variable *NotServed*, about 46% of the observed third-parties have not been connected to piracy websites before October 21, 2016. Third-party domains with EU-TLDs account for some 8% of the observations. On average, EU domains seem to be connected to less piracy websites than non-EU domains, and the same applies to placebo websites.

In what follows, we go beyond the simple comparison between EU and non-EU third-party domains, and introduce an econometric specification that will allow us to estimate the effects of the self-regulation policy.

[21] Note that simply parsing the HTML source code for occurrences of third-party plugins is not enough when client-side scripts, which are executed by the browser, make additional requests.

[22] See https://ghostery.zendesk.com/hc/en-us/articles/115000740394-What-are-the-new-tracker-categories-.

[23] Our results are not sensitive to this specific measure, see the robustness checks detailed in section 4.3.

[24] We only count the appearance of a specific third-party domain once per piracy/placebo domain even if we observe multiple HTML pages per piracy/placebo domain.

3.3 Identification strategy

Reduced form causal evidence on the effects of the recent implementation of the "follow the money" enforcement approach can be generated in a difference-in-differences analysis. The econometric model is based on a before and after comparison of the difference between the treatment (EU=1) and the control group (EU=0), i.e.

$$Log(NumberWebsites_{it}^{k}+1)=\alpha+\delta\,(After_t\times EU_i)+\gamma\,(t\times EU_i)+\nu_t+\mu_i+\varepsilon_{it}, \quad (1)$$

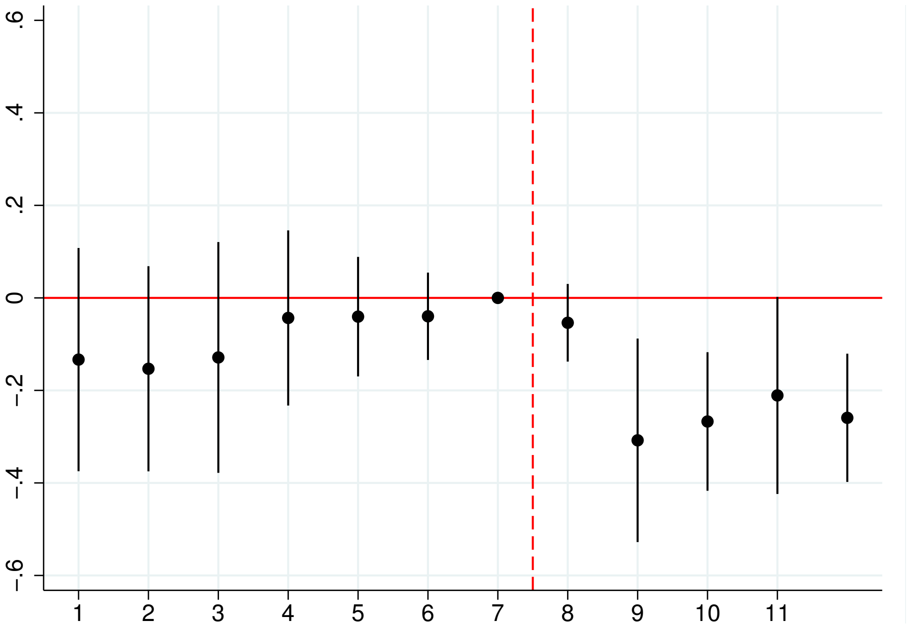
where $NumberWebsites_{it}^{k}$ is the number of websites of type k (piracy, placebo) served by third-party service i at time t. $After_t$ indicates the time period after the agreement to self-regulate came into effect on October 21, 2016. The preferred specification further includes a group-specific linear time trend γ, time fixed effects ν_t and third-party service fixed effects μ_i. Therefore we cannot separately identify coefficients for EU_i and $After_t$ as they are absorbed by third-party service and time fixed effects, respectively. The error term ε_{it} has the standard assumptions, and we report estimates clustered by third-party service.

One might be concerned about a potential threat to exogeneity of the quasi experiment induced by the EC. The decisions of advertising firms may be endogenous in the sense that they anticipate the agreement and start to deactivate accounts of piracy websites before the actual implementation of the agreement. If this is what is happening, we should see differential trends in the piracy market engagement of EU-based and non-EU-based before October 21, 2016. We can test this using the following specification, where we compare the number of piracy websites served by EU-based and non-EU-based third-party services, snapshot by snapshot:

$$Log(NumberWebsites_{it}^{k}+1)=\alpha+\sum_{\tau=1}^{T}\delta^{\tau}\,(\nu_\tau\times EU_i)+\nu_t+\mu_i+\varepsilon_{it}. \quad (2)$$

If EU-based third-party services adapt their behavior in expectation of the agreement, we would expect significant δ^{τ} coefficients before October 21, 2016. Estimation results, alongside with 90% confidence bands are depicted in Figure 4. The estimates suggest that the number of piracy websites served by EU-based advertising services did not change

Figure 4: Group differences over time



Note: *Vertical axis: OLS coefficients of δ^{τ} coefficients as defined in equation (2). The dependent variable is the log number of piracy websites served per third-party service and week. Bars indicate 90% confidence bands, estimated using standard errors clustered at the third-party service level. Note that we report coefficients of a model that is weighted by the popularity of the advertising service (see the discussion in section 4). The significance of pre-agreement coefficients does not change if we estimate a unweighted model. Horizontal axis: Data snapshots in 2016 and 2017 (see figure 2).*

differently than the number of piracy websites served by non-EU-based advertising services prior to the agreement. This provides supportive evidence in favor of the identifying assumption of the difference-in-differences model, i.e. that treatment (EU) and control group (non-EU) follow similar, statistically non distinguishable, trends before the experiment happens.

4 Results

In the following we first describe the results of an estimation of equation (1), and then extend our model to test for indirect effects of the self-regulation. We discuss the robustness of our estimates with respect to measurement error, and with respect to confounding factors using our placebo websites.

4.1 Baseline effects

Column (1) of Table 2 reports the results of estimating equation 1. The coefficient is negative, as expected, but not significantly different from zero. In column (2) we explicitly allow for heterogeneity by interacting $After$ and $After \times EU$ with our measure of third-party service popularity. The triple interaction is negative and significant. For the average advertising service (the average popularity score of EU-based ad services is 0.21), this implies a reduction in the number of piracy websites of about 4%. As described above, the industry is highly concentrated, such that taking the average is perhaps not most informative. We therefore estimate a weighted model in column (3), explicitly giving more weight to advertising services according to their overall popularity. The resulting coefficient implies that the weighted average effect is about −17%.[25] Figure 5 illustrates these results. The dashed red lines indicate the average effect and the weighted average effect, respectively. The black dots indicate in-sample predictions using the coefficient estimates in column (2) of Table 2, with corresponding 90% confidence bands. The concentrated market structure becomes evident in the fact that there are much more observations with small values of popularity than there are observations with large values of popularity. Simply treating large values as outliers, and perhaps excluding them from the analysis would not adequately reflect their empirical importance.

4.2 Indirect effects

In the following, we explore further dimensions of heterogeneity, distinguishing between incumbents and entrants, and between advertising and non-advertising services. To save space and degrees of freedom, we only report results of weighted regressions, but note that all results hold qualitatively and statistically if we instead estimate models with higher-order interactions.

4.2.1 Strategic response

After we have established that the self-regulation has led to a reduction of the number of piracy websites served by advertising services, we now investigate whether this may have

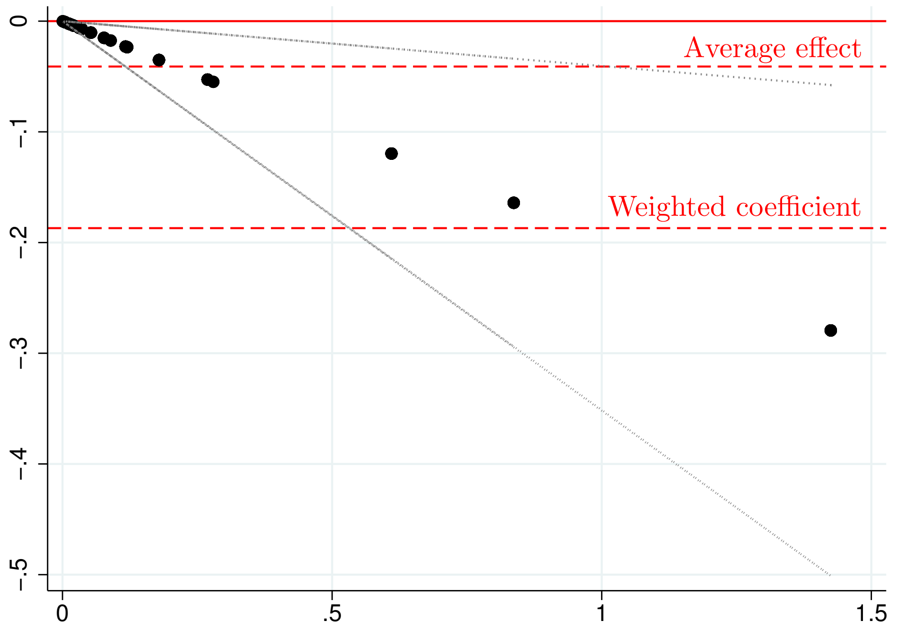
[25]Throughout the paper, we calculate effect sizes as follows: $(exp(Coeff) - 1) * 100$.

Table 2: Baseline effects

	(1)		(2)		(3)	
After × EU	-0.067	(0.058)	-0.023	(0.061)	-0.187**	(0.090)
After × Popularity			-0.000	(0.002)		
After × EU × Popularity			-0.196**	(0.094)		
Weighted by Popularity	No		No		Yes	
Observations	4792		4792		4792	
$\overline{R}^2$	0.965		0.965		0.989	

Dependent variable: Log number of piracy websites served per advertising service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
Popularity indicates the share of websites that use the advertising service.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
Coefficients in column (3) come from a model that is weighted by *Popularity* i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Figure 5: Baseline effects, average and weighting



Note: *The baseline effect as a function of popularity of the advertising service in percent. Dots indicate in-sample predictions based on the $After \times EU \times Popularity$ coefficient in column (2) of Table 2.*
Vertical axis: OLS coefficients. Horizontal axis: Popularity as observed in the estimation sample.
The black dotted lines indicate the 90% confidence band of the in-sample prediction.
The red dashed lines indicate the average effect ($After \times EU \times Popularity * \overline{Popularity}$), and the weighted regression coefficient ($After \times EU$ in column (3) of Table 2).

generated strategic responses. We do this by using variation that comes from the fact that we observe the third-party services used by "placebo" websites, i.e. websites that do

Table 3: Strategic response

	(1)		(2)		(3)	
After × EU	-0.187**	(0.090)			-0.205**	(0.084)
After × Not Served			0.015	(0.022)	0.013	(0.022)
After × EU × Not Served					0.133*	(0.070)
Weighted	Yes		Yes		Yes	
Observations	4792		4792		4792	
$\overline{R}^2$	0.989		0.989		0.989	

Dependent variable: Log number of piracy websites served per advertising service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
Not Served indicates that the service has not serviced piracy websites before October 21, 2016.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

not provide unlicensed content, but are otherwise similar to (e.g. in terms of audience size) to the piracy websites in our data. We construct an indicator variable $NotServed_i$ to distinguish between advertising services that were connected to at least one piracy website before October 21, 2016 and those that were not. Column (3) of Table 3 reports the results of an estimation that adds interactions with this variable. The coefficient of $After \times EU$ remains negative and significant, $After \times NotServed$ is positive, yet small and not significant, and $After \times EU \times NotServed$ is positive and significant. Note however, that this estimate is not very precise. We interpret this as weak evidence that some advertising services increase their presence on piracy websites due to the self-regulation. Two potential mechanisms can explain this finding. First, balancing the expected costs of the sanctions for not complying to the self-regulation against the gains of serving additional eyeballs, it may be beneficial for some firms to enter the market of piracy websites. Second, after being excluded from certain advertising services, piracy websites may actively search for alternatives and sign up elsewhere. Both mechanisms seem plausible, but we have no way to empirically distinguish between the two.

4.2.2 Advertising versus non-advertising services

So far, our analysis was purely based on advertising services. We now test whether non-advertising services react differently to the self-regulation. We do this by estimating

Table 4: Indirect effects, contrasting advertising and other third-party services

	(1)		(2)		(3)	
After × EU	-0.142*	(0.075)			-0.029	(0.101)
After × Ad-Service			0.005	(0.020)	0.006	(0.020)
After × EU × Ad-Service					-0.156**	(0.070)
Weighted	Yes		Yes		Yes	
Observations	23316		23316		23316	
$\overline{R}^2$	0.994		0.994		0.994	

Dependent variable: Log number of piracy websites served per third-party service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the service-level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

an interacted model on the pooled sample. The results in column (3) show that only advertising services reduce their presence on piracy websites. The coefficient of $After \times EU$, indicating the effect for non-advertising services is small and not significant, where as the coefficient of $After \times EU \times Ad - Service$ is similar in size to our baseline estimates (about -14%) and significantly different from zero. This result is somewhat surprising given that most third-party services which we classify as non-advertising are indirectly connected to the online advertising ecosystem. Hence, one would expect that the self-regulation should also be at least partly binding for these services. However, services that track consumers' web browsing behavior collect data that can be valuable for advertisers. As a result, there can be strong monetary incentives not to move out of the piracy market, potentially stronger than the expected costs of sanctions.

4.3 Robustness checks

4.3.1 Measurement error in the treatment group

One might be worried that there is measurement error in our definition of the treatment group. So far, our definition of EU-based third-party services was based on the idea that EU-based firms tend to use TLDs associated with EU countries. Of course, it is always possible for an EU-based firm to register non-country-specific TLDs, such as *.com*, *.net*, *.org*, etc. This type of measurement error should work against us. To see this, consider the following example. First, assume the true effect is negative. If we incorrectly categorize some EU-based third-party services as non-EU-based (type II error), the estimate of the control group average in the post-period is too small (biased towards the average of the true treatment group), because we incorrectly take treated observations into account. As a result, the difference between treatment and control group averages is too small (biased towards zero). Assume the true effect is positive. Now, the estimate of the control group average in the post-period is too large (biased towards the average of the true treatment group) and the difference between treatment and control group averages is too small (biased towards zero). Note that, following the same logic, our estimates would be biased towards zero also in the case where we incorrectly categorize some non-EU-based third-party services as EU-based (type I error). In addition to this theoretical argument, we carry out an empirical test to check for the robustness of our results regarding the potential measurement issue. We do this by collecting additional information from three different data sources. First, we access the *Whois-Register* to collect information on the country of the registrant of the third-party domain. Second, we use information on the country of the IP address of the server that responds to a ping on the third-party domain. Third, we manually search for the geographic headquarters of the parent group of each third-party domain using company information on *LinkedIn*. We then use the additional information to classify third-party services with non-EU-TLDs as EU-based, wherever the additional information suggests that. The results in columns (2)–(4) of Table A.1 show that the estimated effect remains quantitatively and statistically similar to those of our baseline results in column (1).

4.3.2 Counterfactual and placebo exercises

One might be concerned that our results are simply a reflection of unobserved variation that is somehow correlated with our empirical measures. A number of placebo exercises and counterfactual analyses suggest that this concern is not very likely to be valid.

First, we analyze a placebo experiment where we test the hypothesis that legitimate websites, which are similar in terms of audience size, are not affected by the agreement to self-regulate. In column (1) of Table A.2, we show that EU-based and non-EU-based advertising services do not differ in the number of connected placebo websites. In columns (2) and (3), the same holds true when we distinguish between service that have been active in piracy market before and those that have not. Turning to the test for differences between advertising and non-advertising services in Table A.3, we also find no significant effects with regards to placebo websites. We conclude that we cannot reject the hypothesis that legitimate websites are not affected by the self-regulation agreement.

Second, to test whether our results are driven by some kind of seasonality, we run a counterfactual analysis using data from the same time period, but one year before. Correspondingly, we set the date of the placebo experiment to Oct, 21 2015 and re-run the same specifications as in Tables 2, 3 and 4. The coefficient of $After \times EU$ in column (1) of Table A.4 is small and not significantly different from zero. The same holds true in column (2). We also don't see evidence for an increase in the number of piracy websites connected to EU-based advertising services that have not been active in that market before the placebo experiment ($After \times EU \times Ad - Service$). However, there is evidence for a slight expansion of non-EU based advertising services into the piracy market ($After \times NotServed$ is positive and significant). In column (3), we test for differences between advertising and non-advertising services and don't find any significant coefficients.

5 Discussion

After we have established the direct and indirect effects of the self-regulatory agreement on the number of piracy websites served by different types of third-party services, we will discuss potential mechanisms that can explain the results. We also speculate about the self-regulation's potential impact on overall efficiency, and on its success in reaching its overarching goal: a reduction in the supply of online piracy.

5.1 The firm size effect

Our results suggest that firm size is an important determinant of compliance to a self-regulatory initiative. This is in line with some of the previous literature (see e.g. Arora and Cason, 1995; King and Lenox, 2000). A possible mechanism could be that larger firms have more to lose when the regulator decides to go beyond setting incentives to self-regulate. However, there is also evidence showing that regulation can provide a competitive advantage to larger firms (Thomas, 1990). Larger firms may be more likely to be aware of the self-regulatory effort, and therefore also more likely to participate in the stakeholder meetings that the EC has organized, more likely to have signed the agreement. There is no publicly available information on which firms participated in the stakeholder meetings or have signed the agreement. However, it seems likely that it is mostly larger firms that are members of trade associations. For example, the official press release quotes representatives of major industry associations such as the *Joint Industry Committee for Web Standards in the UK and Ireland* and the *UK Internet Advertising Bureau.*[26] A third reason may be that small firms are more likely to be specialized in the piracy market, and therefore follow a different risk strategy in which the expected costs of forced public regulation are discounted more than in less specialized firms. Indeed, there is a (small) negative correlation between a third party service's overall popularity and how much piracy websites make up of its overall number of connected websites in our data.

5.2 Potential efficiency effects

A review of the theoretical literature on the economics of online advertising suggests that sharing data about consumer preferences across platforms can make the advertising market, at least when legitimate publishes are concerned, more efficient (e.g. Ghosh et al., 2015; D'Annunzio and Russo, 2017). While we have no direct way of testing this, the fact that we do not find that pure tracking services change their presence on piracy websites implies that consumers visiting those websites can still be tracked. Empirical studies have shown that information about a consumer's web browsing history can be valuable to advertisers (Lambrecht and Tucker, 2013; Bleier and Eisenbeiss, 2015). The specific information that a consumer has visited a piracy website may have value, because it can be used for positive and negative targeting. On the one hand, observing that a consumer has downloaded an unlicensed copy of a song, film, or software application can be indicative of that consumer's tastes, and of that consumer's demand for complementary products. That such a rationale is perhaps not too far fetched is illustrated in the example of the

[26] See http://ec.europa.eu/growth/tools-databases/newsroom/cf/itemdetail.cfm?item_id=8974&lang=en.

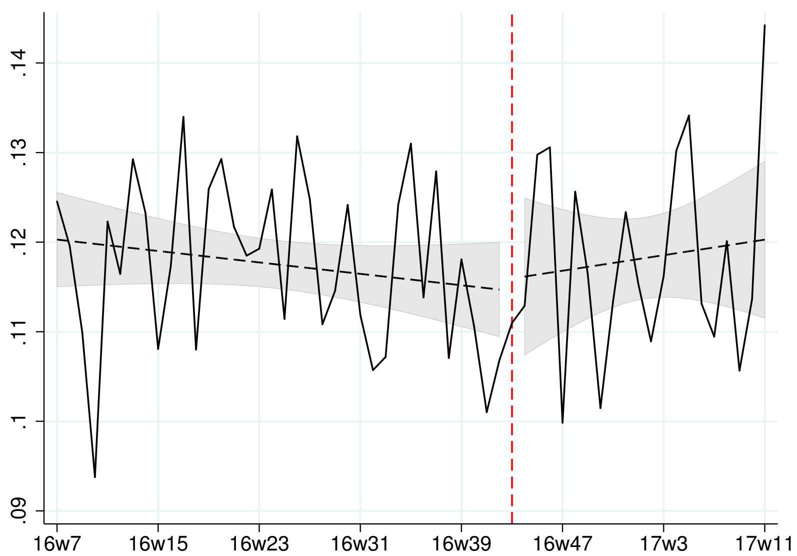
subscription-based licensed video-streaming platform *Netflix*. Reportedly, Netflix examines what's popular on piracy websites to inform their decisions regarding the licensing of new content for their platform.[27] On the other hand, observing that a consumer has visited a piracy website may serve as a signal of that consumer's willingness to pay for specific types of products. Hence, using this information, advertisers can save otherwise wasted ad impressions.

5.3 Overall effectiveness

After we have established the direct and indirect effects of the self-regulatory agreement on the number of piracy websites served by different types of third-party services, we will speculate whether the intervention was successful in reaching its overarching goal: a reduction in the supply of online piracy.

5.3.1 Supply-side perspective

Figure 6: Entry rate



Note: *Vertical axis: Share of domains that appear for the first time in a given week in the Google Copyright Takedown Request data.*
Horizontal axis: Weeks in 2016 and 2017.
Shaded area is the 90% confidence band of a linear regression (dashed line).
Red dashed line indicates the EU self-regulation.

[27] See http://variety.com/2013/digital/news/how-netflix-uses-piracy-to-pick-its-programming-1200611539/.

A necessary condition for a reduction in the supply of online piracy would be that the rate of entry of new piracy websites decreases. To investigate this, we use publicly available data from Google's Transparency Report, which details all take-down requests that copyright owners have filed requesting the removal of infringing content from search results.[28]
Figure 6 depicts the share of domains that appear for the first time in a given week in Google's Transparency Report data. There is no clearly visible downward trend. This is also reflected in a more formal analysis, where we estimate a linear time trend. We plot the corresponding OLS coefficients for the before and after-periods along with 90% confidence bands in Figure 6. The confidence bands overlap, suggesting that there is no significant change in the entry rate of piracy domains before and after the agreement of October 21, 2016. However, it is important to note that this is analysis is purely correlational, because we lack a clear counterfactual.
Speculating even further, it may well be that piracy websites switch their "business model" from selling consumer eyeballs to advertisers to selling consumer data to advertisers.[29] Alternatively, piracy websites may switch to not directly ad-related monetization mechanisms, such as affiliate marketing or subscription models, or even use trackers that take advantage of their visitors' computing power in order to mine cryptocurrencies.[30]

5.3.2 Demand-side perspective

If the self-regulation reduces the revenue stream of piracy websites, and investment in quality is a function of revenues, then this should lead to a decrease in quality. With lower quality, we should also see a reduction in demand. To test this, we collect historical data on the weekly pageviews of the piracy websites in our sample from market research firm *Alexa*. We then aggregate our raw data at the piracy-domain-snapshot-level, and estimate a model similar to equation (1). The dependent variable is the log number of weekly pageviews (in millions). We are interested in testing whether, due to the self regulation,

[28] This data is not optimal for our purposes, because it includes takedown requests for a wide range of copyright-related issues that are not necessarily included in our definition of consumer-focused online piracy. For example, the data also include requests to take down search results based on infringement of product images used in online stores.

[29] For example, a company that offers this type of product to website owners, using a tracking cookie, is *Adthink.com*, see https://big.exchange/sellers.

[30] See http://www.bbc.com/news/technology-41518351.

Table 5: Pageviews of piracy websites

	(1)		(2)		(3)	
After × EU	-0.019	(0.033)	0.067	(0.052)	-0.062	(0.045)
After × Popularity			0.001	(0.001)		
After × EU × Popularity			-0.005**	(0.002)		
Weighted	No		No		Yes	
Observations	3018		3018		3018	
$\overline{R}^2$	0.941		0.941		0.946	

Dependent variable: Log number of weekly pageviews per piracy websites (in millions).
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
Popularity indicates the average share of websites that also use the advertising services that the focal piracy website uses. All models include a group-specific linear time trend, piracy website and time period fixed effects. Coefficients in column (3) come from a model that is weighted by *Popularity*.
Standard errors in parentheses, clustered on the piracy website level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

visits to piracy websites that were connected to at least one EU-based advertising service before the agreement change in a different way compared to visits of piracy websites that were not connected to any EU-based advertising service. The results are reported in Table 5. We do not find much evidence in favor of this mechanism. The coefficient of $After \times EU$ in column (1) is small and not significant. When we allow for heterogeneity in terms of the average size of connected advertising services, we get a significant three-way interaction coefficient in column (2). The estimates imply that the average effect is about -8%.[31] When we estimate a model that is weighted by popularity in column (3), we get a similarly sized effect. However, the estimate is not significantly different from zero. We conclude that this exercise does not provide very convincing evidence that the self-regulation has substantially reduced the demand for online piracy.

6 Conclusions

In this paper, we have evaluated the effectiveness of "follow the money" copyright enforcement based on a private-public self-regulation effort. We use a unique dataset that allows us to observe the interconnections between third-party advertising and consumer tracking services, and piracy websites and legitimate websites. We compare the dynamics before

[31] The mean of popularity is 15.99 in the estimation sample.

and after the advertising industry, under pressure from the European Commission, has agreed on basic principles of an self-regulatory agreement to reduce the financial incentives for piracy websites.

Our results suggests that the presence of advertising services on piracy websites does not change significantly, at least not on average. Once we allow for heterogeneity in terms of size, we show that more popular advertising services, i.e. those that are overall more diffused on the Internet, reduce their presence on piracy websites significantly more. The average advertising services is connected to about 4% less piracy websites, and the weighted average effect implies a reduction of about 17%. We provide some weak evidence that advertising services that were not active in the piracy market before are more likely serve piracy websites after the agreement. Finally, we show that non-advertising services, i.e. services that are primarily in the business of collecting (and selling) data about consumers' web behavior, do not change their presence on piracy websites – irrespective of their overall popularity. In that sense, our paper shows that firms indeed "follow the money", but not not necessarily in the way the self-regulatory effort intended to.

We speculate about the success of the self-regulation regarding its overarching goal to narrow down online piracy using data on entry of new piracy websites and visits to piracy websites, and conclude that there is at least no clear downward trend.

Our study contributes to the literature on self-regulation, the online advertising industry and online copyright enforcement. The results also provide implications for different settings where the availability of online advertising creates incentives to produce content that has negative externalities.

References

Acquisti, A., Taylor, C., and Wagman, L. (2016). "The Economics of Privacy." *Journal of Economic Literature*, *54*(2), 442–492.

Adermon, A., and Liang, C.-Y. (2014). "Piracy and Music Sales: The Effects of an Anti-Piracy Law." *Journal of Economic Behavior & Organization*, *105*, 90–106.

Aguiar, L., Claussen, J., and Peukert, C. (2017). "Catch Me If You Can: Effectiveness and Consequences of Online Copyright Enforcement." *Working paper.*

Allcott, H., and Gentzkow, M. (2017). "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives*, *31*(2), 211–236.

Arora, S., and Cason, T. N. (1995). "An Experiment in Voluntary Environmental Regulation: Participation in EPA's 33/50 Program." *Journal of Environmental Economics and Management*, *28*(3), 271 – 286.

Athey, S., Calvano, E., and Gans, J. (forthcoming). "The Impact of Consumer Multi-homing on Advertising Markets and Media Competition." *Management Science.*

Bae, S. H., and Choi, J. P. (2006). "A Model of Piracy." *Information Economics and Policy*, *18*(3), 303–320.

Bleier, A., and Eisenbeiss, M. (2015). "Personalized Online Advertising Effectiveness: The Interplay of What, When and Where." *Marketing Science*, *34*(5), 669–688.

Chaabane, A., Kaafar, M., and Borelli, R. (2012). "Big Friend is Watching You: Analyzing Online Social Network Tracking Capabilities." *WOSN 12 Proceedings.*

Danaher, B., and Smith, M. (2014). "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales." *International Journal of Industrial Organization*, *33*, 1–8.

Danaher, B., Smith, M. D., Telang, R., and Chen, S. (2014). "The Effect of Graduated Response Anti-Piracy Laws on Music Sales: Evidence from an Event Study in France." *Journal of Industrial Economics*, *62*(3), 541–553.

D'Annunzio, A., and Russo, A. (2017). "Ad Networks, Consumer Tracking, and Privacy." *Working Paper.*

de Cornière, A., and de Nijs, R. (2016). "Online Advertising and Privacy." *RAND Journal of Economics*, *15*(3), 311–327.

Egorov, G., and Harstad, B. (2017). "Private Politics and Public Regulation." *Review of Economic Studies*, *84*, 1652–1682.

Ellman, M., and Germano, F. (2009). "What Do The Papers Sell? A Model of Advertising and Media Bias." *Economic Journal*, *119*, 680–704.

Ghosh, A., Mahdian, M., McAfee, R. P., and Vassilvitskii, S. (2015). "To Match or Not to Match: Economics of Cookie Matching in Online Advertising." *ACM Transactions on Economics and Computation*, *3*(2), article 12.

Gill, P., Erramili, V., Chaintreau, A., Krishnamurthy, B., and Papagiannaki, D. (2013). "Follow the Money ? Understanding Economics of Online Aggregation and Advertising." *Proceedings of the 2013 ACM Internet Measurement Conference.*

Goldfarb, A. (2014). "What is Different About Online Advertising?" *Review of Industrial Organization*, *44*, 115–129.

Hendel, I., Lach, S., and Spiegel, Y. (2017). "Consumers' Activism: The Cottage Cheese Boycott." *RAND Journal of Economics*, *48*(4), 972–1003.

King, A., and Lenox, M. (2000). "Industry Self-Regulation without Sanctions: The Chemical Industry's Responsible Care Program." *Academy of Management Journal*, *43*(4), 698–716.

Lambrecht, A., and Tucker, C. (2013). "When Does Retargeting Work? Information Specificity in Online Advertising." *Journal of Marketing Research*, *50*(10), 561–576.

Lerner, A., Kornfeld Simpson, A., Kohno, T., and Roesner, F. (2016). "Internet Jones and the Raiders of the Lost Tracker: An Archaeological Study of Web Tracking from 1996 to 2016." *Proceedings of the 25th USENIX Security Symposium.*

Libert, T. (2015). "Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on One Million Websites." *International Journal of Communication*, *9*, 3544–3561.

Maxwell, J., Lyon, T., and Hackett, S. (2000). "Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism." *Journal of Law & Economics*, *43*(10), 583–617.

McKenzie, J. (2017). "Graduated Response Policies to Digital Piracy: Do They Increase Box Office Revenues of Movies?" *Information Economics and Policy*, *38*(1), 1–11.

Peitz, M., and Reisinger, M. (2015). "The Economics of Internet Media." In S. Anderson, J. Waldfogel, and D. Stromberg (Eds.), *Handbook of Media Economics*, vol. 1B, 445–530, Elsevier.

Peukert, C., Claussen, J., and Kretschmer, T. (2017). "Piracy and Box Office Movie Revenues: Evidence from Megaupload." *International Journal of Industrial Organization*, *52*, 188–215.

Poort, J., Leenheer, J., van der Ham, J., and Dumitru, C. (2014). "Baywatch: Two approaches to measure the effects of blocking access to The Pirate Bay." *Telecommunications Policy*, *38*, 383–392.

Reimers, I. (2016). "Can Private Copyright Protection be Effective? Evidence from Book Publishing." *Journal of Law and Economics*, *59*(2), 411–440.

Roosendaal, A. (2012). "We Are All Connected to Facebook... by Facebook!" In S. Gutwirth, R. Leenes, P. DeHert, and Y. Poullet (Eds.), *European Data Protection: In Good Health?*, 3–19, Springer, New York.

Schelter, S., and Kunegis, J. (2016). "On the Ubiquity of Web Tracking: Insights from a Billion-Page Web Crawl." *arXiv:1607.07403.*

Thomas, L. (1990). “Regulation and Firm Size: FDA Impacts on Innovation.” *RAND Journal of Economics*, *21*(4), 497–517.

Tucker, C. (2012). “The Economics of Advertising and Privacy.” *International Journal of Industrial Organization*, *30*, 326–329.

Tucker, C. (2014). “Social Networks, Personalized Advertising and Privacy Controls.” *Journal of Marketing Research*, *51*(10), 546–562.

Xiang, Y., and Sarvary, M. (2007). “News Consumption and Media Bias.” *Marketing Science*, *26*(5), 611–628.

Table A.1: Direct effects, different measures of *EU-based*

	(1)	(2)	(3)	(4)
After × EU	-0.187**			
	(0.090)			
After × EU (Reg)		-0.148***		
		(0.051)		
After × EU (IP)			-0.114**	
			(0.053)	
After × EU (LinkedIn)				-0.087*
				(0.046)
Weighted	Yes	Yes	Yes	Yes
Observations	4792	4792	4792	4792
$\overline{R}^2$	0.989	0.989	0.989	0.989

Dependent variable: Log number of piracy websites served per third-party service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country. *EU (Reg)* indicates whether the third-party domain is an EU-country TLD or is registered in an EU-country. *EU (IP)* indicates whether the third-party domain is an EU-country TLD or the responding IP address is located in an EU-country. *EU (LinkedIn)* indicates whether the third-party domain is an EU-country TLD or the corresponding company is headquartered in an EU-country (according to LinkedIn).
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Table A.2: Placebo: Baseline effect and entry

	(1)		(2)		(3)	
After × EU	0.166	(0.154)			0.171	(0.156)
After × Not Served			0.019	(0.080)	0.019	(0.082)
After × EU × Not Served					-0.043	(0.098)
Weighted	Yes		Yes		Yes	
Observations	4792		4792		4792	
$\overline{R}^2$	0.976		0.976		0.976	

Dependent variable: Log number of placebo websites served per third-party service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
Not Served indicates that the service has not serviced piracy websites before October 21, 2016.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Table A.3: Placebo: Indirect effects, contrasting ad and non-ad services

	(1)		(2)		(3)	
After × EU	0.165	(0.118)			0.178	(0.119)
After × Ad-Service			0.014	(0.023)	0.014	(0.023)
After × EU × Ad-Service					-0.026	(0.060)
Weighted	Yes		Yes		Yes	
Observations	23316		23316		23316	
$\overline{R}^2$	0.990		0.990		0.990	

Dependent variable: Log number of placebo websites served per third-party service.
After indicates period after October 21, 2016, *EU* indicates a TLD of a EU-country.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the service-level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Table A.4: Counterfactual: Direct and indirect effects, one year before

	(1)		(2)		(3)	
After × EU	-0.048	(0.059)	-0.044	(0.083)	-0.055	(0.079)
After × Not Served			0.046***	(0.015)		
After × EU × Not Served			-0.049	(0.050)		
After × Ad-Service					-0.039	(0.025)
After × EU × Ad-Service					0.032	(0.066)
Weighted	Yes		Yes		Yes	
Observations	4593		4593		19905	
$\overline{R}^2$	0.990		0.990		0.994	

Dependent variable: Log number of piracy websites served per third-party service.
After indicates period after October 21, 2015, *EU* indicates a TLD of a EU-country.
Not Served indicates that the service has not serviced piracy websites before October 21, 2015.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
All models are weighted by *Popularity*, i.e. the share of websites that use the third-party service.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$