Follow the Money
Online Piracy and Self-regulation in the Advertising Industry

Batikas, Michail; Claussen, Jörg; Peukert, Christian

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

Michail Batikas, Jörg Claussen, Christian Peukert

Accepted Manuscript

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

Michail Batikas, Jörg Claussen, Christian Peukert

Please cite this article as: Michail Batikas, Jörg Claussen, Christian Peukert, Follow The Money: Online Piracy and Self-Regulation in the Advertising Industry, *International Journal of Industrial Organization* (2019), doi: https://doi.org/10.1016/j.ijindorg.2019.02.001

Highlights

- We study the effects of an EU self-regulation effort in the online advertising industry

- The aim is to reduce advertising revenues for piracy websites

- The effort is limited in its effectiveness

- Only the largest EU-based advertising services comply

- No evidence that demand for piracy websites decreases

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

Michail Batikas[1], Jörg Claussen[1,3], and
Christian Peukert[2]

[1]LMU Munich
[2]UCP – Católica Lisbon School of Business and Economics
[3]Copenhagen Business School

Abstract

We study the effects of a self-regulation effort, orchestrated by the European Commission in 2016 and finalized in 2018, that aims to reduce advertising revenues for publishers of copyright infringing content. Data on the third-party HTTP requests made by a large number of piracy websites lets us observe the relations of the piracy and advertising industry over time. We compare these dynamics to a control group of non-advertising services which are not subject to the self-regulation. Our results suggest that the effort is limited in its effectiveness. On average, the number of piracy websites that make requests to EU-based advertising services does not change significantly. Only when we allow for heterogeneity in the popularity of third-party services, we find that the number of piracy websites that interact with the most popular EU-based advertising services decreases by 42%. We do not find evidence that non-EU-based advertising services react to the self-regulation. This implies that only a small share of the firms in the market comply with self-regulation in a way that is visible in our data. We also do not find evidence that the demand for piracy websites decreases due to this "follow the money" initiative.

Keywords: Piracy, Copyright Enforcement, Online Advertising, Natural Experiment
JEL No.: K4, L5, L8

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 provision of services. The classic example is the newspaper market, where advertisers fund journalism by effectively subsidizing the price charged to readers. However, these incentives may be less socially desirable if the provided services are illegal, as it is the case for the provision of unlicensed software, music, film, and book content ("online piracy"). According to industry reports, copyright infringing websites are predominately funded by advertising.[1] When advertising makes it profitable to sustain a platform that makes unlicensed software and media products available to consumers, and consumers substitute away from licensed products, this reduces the profits of creators. While close to zero prices increase consumer surplus in the short run, welfare can be reduced in the long run if profits are too low for continued investment in new products (e.g. Bae and Choi, 2006).

Recognizing this trade-off, firms and governmental authorities have used a variety of different approaches to online copyright enforcement. A growing body of literature evaluates these policies and arrives at sobering conclusions. The effects of anti-piracy interventions are often short-lived and have collectively failed to substantially eradicate online piracy (Danaher et al., 2014; Adermon and Liang, 2014; Reimers, 2016; McKenzie, 2017; Peukert et al., 2017; Aguiar et al., 2018).[2] An important reason why new incarnations of unlicensed content platforms keep entering the market could be that the previous initiatives did not put enough emphasis on the economic incentives to supply unlicensed content. A relatively new approach to online copyright enforcement, termed "follow the money", aims to do so. The largest effort of this kind is an initiative orchestrated by the European Commission (EC), in which the online advertising industry faces pressure to establish a voluntary agreement to reduce the flow of revenues to copyright infringing websites. Following the example of a series of similar smaller initiatives on the national-level, the EC started to invite stakeholders across Europe to participate in a supranational self-regulatory effort. Although participation is voluntary and sanctions for non-compliance are relatively soft in the short-run, the EC has underlined that it has significant leverage in the long run, threatening that soft law measures "[...] could be backed by legislation, if required, to

---

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

[2]For example, the piracy rate in the UK is still 25%. See Intellectual Property Office (2016), "Online Copyright Infringement Tracker, Latest wave of research Mar 16–May 16, Overview and key findings", p. 19.

ensure their full effectiveness."[3] The implementation of the self-regulation proceeded in two steps. In October 2016, participants approved a statement of *Guiding Principles* which include commitments to concrete measures. In June 2018, participants declared publicly to adhere to a perhaps stricter and more binding version of these principles by signing a *Memorandum of Understanding* (MoU).

We evaluate the effects of this self-regulation with respect to observable interactions of piracy websites and online advertising intermediaries, and with respect to the demand for piracy websites. Specifically, we compare the involvement of EU-based and non-EU-based advertising firms in the piracy market, before and after the *Guiding Principles* and the *Memorandum of Understanding*, to a control group of other third-party services that are not directly part of the advertising industry and are therefore not subject to the EC's effort to self-regulation.

We collect data on the advertising and non-advertising services associated with 392 piracy websites and 784 comparable legitimate ("placebo") websites over a period of 3 years, based on the HTTP requests that these websites make to third parties. This lets us follow 7,189 third-party services and their interactions with piracy and placebo websites. We add meta information that lets us distinguish between EU-based and non-EU-based, smaller and larger, and different types of third-party 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 the average effect of the *Guiding Principles* is not distinguishable from zero. Second, once we allow for heterogeneity in size, we find that the number of piracy websites that interact with the largest EU-based advertising services decreases by 42%. We interpret this as evidence that the likelihood of joining the initiative or complying with the self-regulation increases with firm size. Although data issues preclude a clean causal interpretation of the estimates, the *Memorandum of Understanding* seems to have similar effects. 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 no decrease in the number of visits to piracy websites, comparing those that are connected to affected and those that are connected to unaffected third-party services. Being the first academic study that evaluates the "follow the money" approach to online copyright enforcement, our paper contributes to the growing literature on piracy, counterfeiting and the evalua-

---

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

tion of IP enforcement policies (Adermon and Liang, 2014; Danaher and Smith, 2014; Reimers, 2016; Peukert et al., 2017; Aguiar et al., 2018). In the sense that we study a potential source of unintended externalities of online advertising, and an regulatory effort that is specifically targeted at that industry, we also contribute to the literature on the economics of online advertising (see Tucker, 2012; Goldfarb, 2014; Peitz and Reisinger, 2015 for summaries), linking directly to papers that study the relationship between advertising and content quality (Sun and Zhu, 2013; Shiller et al., 2018).

We believe that our results allow to draw some broader implications. Showing that creating incentives for firms to self-regulate and internalize an externality can be intricate when firms are heterogenous, can be helpful for the design of future efforts. Closely related, the EC has announced that it will try to establish similar memoranda in the market for online payment services, and logistics and transport services. Another example is the market for false and misleading information ("fake news"), which seems to share many characteristics with the market for online piracy.

2   Context

2.1   The market for online advertising and consumer tracking

A growing literature on the economics of online advertising has started to analyze the industrial organization of the industry. For example, theoretical work has studied the incentives of publishers to collect and share information about consumers (de Cornière and de Nijs, 2016; Athey et al., 2018), and the welfare implications of consumer opt-out policies (D'Annunzio and Russo, 2017). Few papers study the externalities that online advertising can have on the incentive to supply content, suggesting that advertising is positively related to content quality (Sun and Zhu, 2013; Shiller et al., 2018).

Empirically, the online advertising and consumer tracking industry is comprised of a complex system of actors 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

Figure 1: The online advertising ecosystem
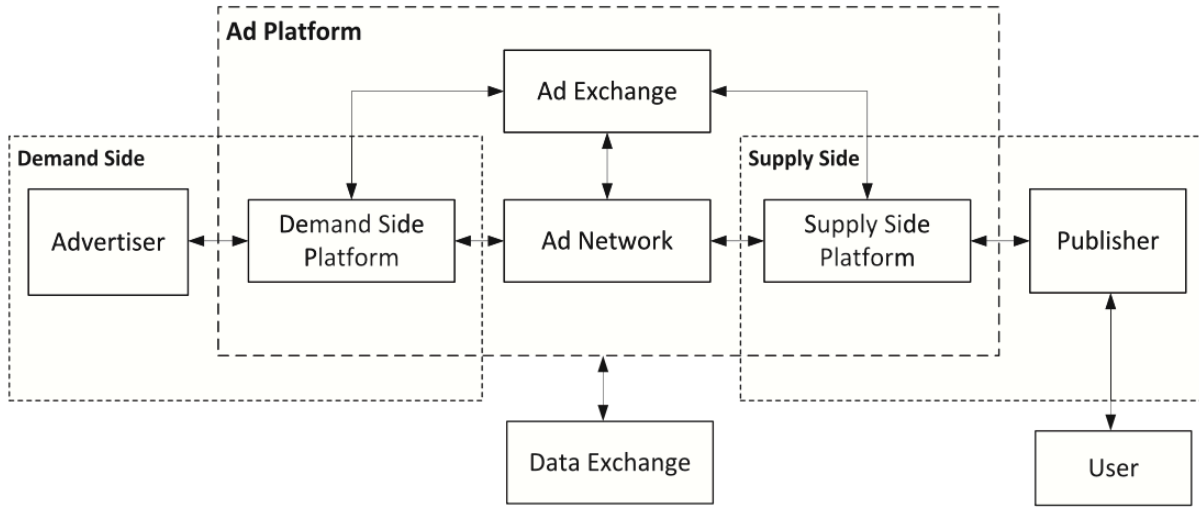
Source: Estrada-Jiménez et al. (2017)

concentrated with the top 20 services covering about a third of the market (Gill et al., 2013; Schelter and Kunegis, 2018).

Figure 1 illustrates some of the key actors and their interrelations in the online advertising ecosystem. *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 processed and made available 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.[4] It is important to note that firms are often

---

[4]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/.

4

Figure 2: Example of third-party requests
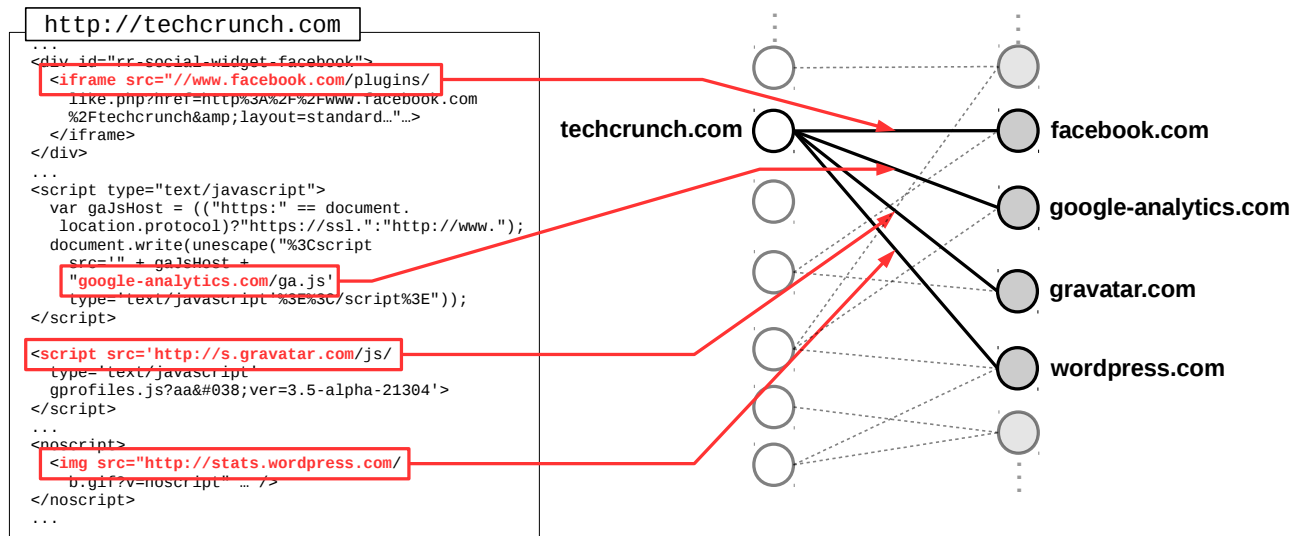
Note: *Similar to Figure 1 in* Schelter and Kunegis (2018).

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.

Many of the interactions between various parties that happen in real-time as a user visits a publisher's website are not directly visible for the consumer, nor are the identities of the involved parties that collect and trade personal information completely transparent. Monitoring a website's third-party HTTP requests can make some of the interactions of websites and the advertising industry observable.[5] Figure 2 illustrates a stylized example of third-party requests: the page hosted on `techcrunch.com` makes requests to third-party domains such as `facebook.com` and `google-analytics.com`, embedding content that is hosted on third-party servers, such as images and scripts. In this way, a website can display content that is controlled by other parties, and this content can differ for each individual request and each individual user. The best example of such content is of course advertisements. However, a website can also embed contents from third parties that are not, or not directly, related to advertising. For example, a website may embed externally hosted video content or outsource certain interactive functionalities, like comment or chat systems, to a specialized third-party provider.

Historical information on the third-party requests of websites is publicly available, and can be matched

---

[5]In fact, several software tools, which come as extensions to the browser, are available to help consumers visualize and understand this data. Some tools, like *Ghostery*, *Privacy Badger*, and *Disconnect* offer the ability to block requests to specific third-parties.

with other publicly available sources of meta information on third-party services. In the empirical analysis below, we will use these datasets in the specific context of online copyright enforcement.

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

2.2.1 The broader regulatory context

In the last decade, firms and governmental authorities have used a variety of different approaches to enforce copyright on the Internet. A growing body of literature evaluates these policies and arrives at sobering conclusions. Stricter private enforcement and public anti-piracy legislation, shutting down and blocking access to certain websites did only lead to relatively small and short-lived decreases in piracy activities (Danaher et al., 2014; Adermon and Liang, 2014; McKenzie, 2017; Poort et al., 2014; Reimers, 2016; Peukert et al., 2017; Aguiar et al., 2018). 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.[6]

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,[7] and annual ad revenues of piracy sites are in the range of $230 million, with profit margins of 80%.[8] Commonly referred to as "follow the money", initiatives have been started in a variety of countries since 2014.[9] In the United Kingdom, a private-public consortium of copyright owners and

---

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

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

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

[9]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/file/d/1ljl8CGG6tSzDpBoOLCv3SVwA-uC7f7Hq.

the City of London Police's Intellectual Property Crime Unit has started to provide a list of websites for advertisers to avoid.[10] 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.[11]

Also on the supranational-level, the EC seems to believe that advertising funded IP infringement is a problem that is more effectively solved by soft law, rather than legislation. The idea seems to be that soft law measures, such as the "follow the money" initiatives, provide context-specific regulatory flexibility by allowing stakeholders to use tools that they have at their discretion, and not whatever is defined in the legislation and perhaps incompatible with certain business models. Consequently, the EC 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.[12] The intention was further underlined in a communication from December 2015, where the EC 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."[13] The EC sees its supranational effort as complementary to national efforts, in that it creates a stimulus for the creation or development of similar national initiatives, where it hopes that more far going commitments can be achieved. Further, although not directly linked, soft law initiatives are complementary to the EC's legislative proposals concerning IP, such as its recent proposal to reform copyright rules. In addition to its effort to establish a self-regulation in the context of online copyright infringement, online advertisers and advertising service intermediaries, the EC has identified three other areas in which it sees scope for intermediary service providers and rights-holders to cooperate in the creation and development of self-regulation: (1) sales of counterfeit products over the internet, (2) online payment service providers, and (3) logistics and transportation service providers. Regarding internet sales platforms, there is already a MoU in place since May 2011, with two reports in 2013 and 2017 suggesting its positive impact on removing counterfeit products

---

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

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

[12]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.

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

from online marketplaces.[14] There is not yet any public information regarding prospective memoranda concerning online payment service providers, and logistics and transportation service providers. The evaluation of the "follow the money" initiative in the context of online advertising is at the core of this paper. In the next section, we lay out the institutional details leading up the MoU in this area.

2.2.2 "Follow the money" and online piracy on the European Level

The rollout of the EC's "follow the money" initiative is characterized by two major milestones. The first is an agreement to *Guiding Principles*, which act as a framework for the negotiations of a final *Memorandum of Understanding*. The latter can be interpreted as a more carefully crafted and potentially more binding version of the former. In what follows, we describe each in more detail.

Agreement to Guiding Principles on October 21, 2016 In the implementation of a soft law measure, the EC can be seen as a honest broker that manages the process of defining the cornerstones of the self-regulation. The first step involved a call for interest sent to organizations and stakeholder groups in the IP and advertising field across Europe, including user organizations and international organizations. These invitations were intended to maximize the number of participants but did not follow a specific structure, such as selecting specific firms by a measure of their market share or revenues. Industry associations were part of the initial set of recipients, also because the EC saw this as way to efficiently reach a large number of individual member organizations. A non-disclosed share of the approached organizations showed interest to dedicate resources to participate in discussions and negotiations. A first meeting was held on March 14, 2016 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."[15] On October 21, 2016, a collective declaration was signed, formalizing the *Guiding Principles* under which the participants agreed to work to on a MoU.[16] 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 document can be interpreted as a framework that reflects the values

---

[14]See https://ec.europa.eu/growth/industry/intellectual-property/enforcement/memorandum-understanding-sale-counterfeit-goods-internet_en.

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

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

that the involved parties believe should lead the direction towards the cooperation, but still includes strong commitments from all sides.[17] These commitments appear to be relatively strong given that the *Guiding Principles* are essentially not more than an industry declaration. However, it is important to note that the identities of the signatories of the *Guiding Principles* have not been publicly disclosed, and that the document does not mention any sanctioning mechanisms in case of non-compliance.

Agreement to a Memorandum of Understanding on June 25, 2018 After a number of subsequent meetings, the final version of the MoU was negotiated and eventually signed on June 25, 2018. While the commitments that were already outlined in the *Guiding Principles* were largely carried over to the MoU, there are also additional aspects, such that the level of commitments is much higher with respect to individual signatories and industry associations. In contrast to the *Guiding Principles*, the *Memorandum of Understanding* also has explicit sanctions. While there are no direct individual sanctions for non-compliance (the EC doesn't have legal competence to impose tangible sanctions), article 26 describes the procedure in which other signatories, after a plenary meeting, can invite a non-complying signatory to withdraw from the MoU. This deterring mechanism may appear quite soft, but it is possible that this is the result of strategic considerations. The prospect of sanctions, which can be circumvented simply by not signing the MoU, can certainly reduce the incentives to take part in the process in the first place. Indeed, it seems that the main goal of the EC is to maximize the number of potential signatories, perhaps to make them aware of the issue of advertising-funded IP infringement in the first place. Despite the threat that the EC may back soft law measures with legislation, the perhaps most important deterring mechanism, however, is that the signatories of the MoU declare publicly that they adhere to the agreement and commit to implement specific obligations.[18]

Given this stark difference, it is not necessarily clear that the signatories of the *Guiding Principles* are the same as the signatories of the MoU. It is not publicly disclosed how many signatories of the *Guiding Principles* left the process, but it seems possible that the number is not zero. The publication of the

---

[17] For example, article 4 is especially explicit: "Under the agreement, signatories directly involved in buying, selling or brokering the sale of advertising space will undertake to include certain safeguards in their contractual agreements, for example but not limited to the use of Content Verification (CV) tools, Ad Delivery and Ad Reporting systems, schedules, online rights monitoring and brand protection services, with the aim that the advertising placed by them (in the case of advertisers) or through their services (in the case of advertising intermediaries) is not associated with commercial scale IP infringing content. Signatories agree to pass along the obligations through the chain of contracts."

[18] In addition to the list of signatories, the EC has also published a video of the signing procedure and some statements of individual signatories. See https://ec.europa.eu/growth/industry/intellectual-property/enforcement/memorandum-of-understanding-online-advertising-ipr_en.

Guiding Principles may have also made additional organizations aware of the initiative. Indirectly, some organizations may have decided to join the process independently after they became aware of it as members of industry associations.[19]

3 Identification strategy and data

In this section, we first introduce the identification strategy and model specification. Then, we describe the multi-stage data collection process and provide some descriptive statistics. Finally, we report some results of a test of a necessary condition of the identifying assumption.

3.1 Identification strategy

Our empirical analysis 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 either agreement, on October 21, 2016 or June 25, 2018. We follow the standard in the literature of copyright enforcement policy evaluation and estimate a simple reduced form model to arrive at the causal effects of the self-regulation effort. Our econometric model is essentially a difference-in-differences analysis that is based on a before and after comparison of the difference between a control group and two treatment groups. We define two separate treatment groups, because we want to be able to distinguish between directly affected firms (i.e. firms that operate within the European Union) and potentially indirectly affected firms (i.e. firms that operate outside of the European Union). This gives us the flexibility to allow for strategic reactions of firms in response to the self-regulation effort. The control group comprises of third-party requests that are not directly related to advertising, and therefore neither subject to the self-regulatory effort itself nor to the threat of legislation.

We begin by estimating the effect of the *Guiding Principles* agreement on the number of piracy websites that make requests to a third-party advertising service. To do so, we formulate the following

---

[19]This would be in line with article 3 of the *Guiding Principles*, which explicitly says that industry associations "will commit to investing their best efforts to encourage their members, in respective sectors, to not offer the sale, recommend, or buy media space on commercial-scale IP infringing sites."

model:

$$Log(NumberWebsites_{it}^{k} + 1) = \alpha + \delta_1 \left(After_t \times EU\text{-}Ad_i \right) + \delta_2 \left(After_t \times Non\text{-}EU\text{-}Ad_i \right)$$
$$+ \gamma EU_i + w_t + \mu_i + \varepsilon_{it}, \tag{1}$$

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 *Guiding Principles* of self-regulation were agreed upon on October 21, 2016. The preferred specification further includes a group-specific linear time trend γ, time fixed effects w_t and third-party service fixed effects μ_i. Therefore we cannot separately identify coefficients for $Non\text{-}EU\text{-}Ad_i$, $EU\text{-}Ad_i$ 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. We later augment this model to test whether the *Memorandum of Understanding* has had additional effects.

The identifying assumption, i.e. that – in the absence of the treatment – the treatment group(s) would have developed over time just as the control group does can of course not be tested directly. However, we can test a necessary condition of this assumption: whether the treatment groups' time trend is parallel to the control group's time trend before the treatment. We report the results of such a test in section 4.2, along with results of further counterfactual analyses in section 4.3 that support a causal interpretation of our baseline results. Before doing so, we first introduce and describe the data set.

3.2 Data collection

In the 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 736 distinct piracy domains. We additionally construct a sample of legitimate websites that reach a similarly sized audience as the piracy websites in our list. Using the ranking of the top 1 million most frequently visited websites provided by market research firm *Alexa*, for each piracy domain, we look for two "placebo" domains that are placed one position above or one position below the piracy domains from our initial list of piracy domains.

---

[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.

In the second step, we collect historical information about the HTTP requests that these websites make to third parties. We use data from *HTTPArchive*, a project that periodically crawls the homepages of about half a million domains. With a few exceptions, these data are available on a biweekly basis, stemming from a crawl at the beginning and in the middle of the month.[21] We have access to data from July 2016 (three months prior to the *Guiding Principles* agreement) to October 2018.[22] This long time structure will help us to investigate the dynamic effects of the self-regulatory efforts. Because three snapshots cover an atypically small amount of websites, we exclude them from our sample.[23] Before we match this data set against our list of piracy websites, we construct a time-varying measure of each third-party domain's overall popularity, across publishers of different types. We do so by counting the total number of websites on which a third-party domain appears in each snapshot of the *HTTPArchive* data. This measure varies from 1 to 417183, with a mean of 2074 (median of 97) and a standard deviation of 13876.

In the third step, we obtain meta information about the third-party domains in the *HTTPArchive* from the market research firms *whotracks.me* and *Evidon*. This allows us to match domain names to company names, membership in industry associations, and most importantly, this data lets us categorize third-party domains according to service types. To define the treatment and control group, we can distinguish between advertising-related and non-advertising-related third party requests (e.g. advertising, site analytics, customer interaction, content delivery, etc.). Around 49.5% 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 sites".[24] 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.[25]

After matching all the available information to create a joint data set, we observe information about 392 piracy websites and 784 placebo websites and their corresponding third-party services.

---

[21] The public dataset, available in Google's BigQuery web service, does not contain information for 12-01-2016, 01-01-2017, 01-15-2017, 04-15-2018, and 05-01-2018.

[22] The earliest snapshot is 07-15-2016 and the latest snapshot is 10-01-2018.

[23] These snapshots are 02-01-17, 02-15-17, and 03-01-17. However, as we show in table A.5, not excluding those observations yields qualitatively similar results. Because of similar issues of missing data, we use a different data source for historical data from 2015, *CommonCrawl*, for a robustness check in section 4.3.

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

[25] Our results are not sensitive to this specific measure, see the robustness checks detailed in section 4.4.2.

Finally, we aggregate our dataset to the third-party-domain-snapshot-level, by counting on how many piracy domains (or how many placebo domains) we observe a given third-party domain at least once in a given time snapshot. In this final data set, we observe 487 advertising services that are not based in the European Union, 43 advertising services that are based in the European Union, and 6659 non-advertising services.

3.3 Descriptive statistics

We plot the relative frequency of third-party services in our main estimation sample in figure A.1. The left-hand side panel plots the distribution of categories across. The category of a majority of 72.2% of third-party domains is not categorized by *whotracks.me* and *Evidon*. Advertising services account for 15.6% of the third-party domains, tracking services (analytics tools and social media buttons) represent 6.7%, and interactive tools (comments, audio/video players, etc.) and hosting/content delivery networks account for less than 5% of the observations. The right-hand side panel, depicting the average number of websites served by each category, suggests that the popularity distribution is very different. The most used category is hosting/CDN (requests from an average of 20,500 websites), followed by advertising and tracking, which each on average serve about 30% of the number of websites served by hosting/CDN services. The average interactive tool is used by about 3750 websites, and the average "unknown" third-party service is requested by less than 250 websites. While striking, our robustness checks in section 4.4.1 suggest that this variation has very little impact on our baseline estimates.

Descriptive statistics by treatment/control group and pre/post period are reported in table A.1. We first see that the average number of websites that send requests to a given third-party service, a measure of popularity or firm size, differs across types. *Non-EU-Ad* services serve an average of 6558 websites in the pre-period, 6892 websites in the post-period. *EU-Ad* services are connected to an average of 1228 and 1340 in the pre- and post-period, respectively. The average number of websites that send requests non-advertising services, i.e. services in the *control* group, is 1220 in the pre-period and 1338 in the post period.

Regarding the variable of most interest, *NumberPiracyWebsites*, we observe a decrease in the post-period compared to the pre-period across all types of third-party services. *Non-EU-Ad* services move from an average of 5.37 to 5.03, *EU-Ad* services from 2.09 to 1.84, and the *control* group from 1.47

to 1.32. The simplest difference-in-difference estimate would therefore suggest that the number of piracy websites on *Non-EU-Ad* services is 3.5% lower than the pre-period mean, and it is 4.7% lower regarding *EU-Ad* services and their pre-period mean. In what follows we extend this (certainly too) simple analysis by estimating the parameters of equation (1) using OLS.

4 Results

In the following we first report the results of our evaluation of the *Guiding Principles* agreement. We estimate equation (1), and put specific emphasis on heterogeneity with regards to firms size and firm type. We discuss the robustness of our estimates with respect to longer time windows, different types of measurement error, and regarding confounding factors using two placebo set-ups. We then turn to an evaluation of the *Memorandum of Understanding*.

4.1 Baseline results: The "Guiding Principles" agreement

Column (1) of Table 1 reports the results of estimating equation (1). The coefficient of *After × EU Ad* is negative, as expected, but not significantly different from zero. The coefficient of *After × non-EU Ad* is negative and not significant as well, but around 80% smaller in terms of magnitude. This is in line with the idea that any reaction to the self-regulation should be stronger for directly affected firms.

The fact that we don't find a significant effect when we look at the response of the average firm can have many reasons. First, joining the self-regulatory process or complying to it, can be against the strategic interests in the competition with other firms. For example, some firms might specialize in serving the piracy market, or some firms might immediately start to serve the piracy market once another firm has exited it. Note that the latter process must not necessarily be driven by active involvement of the advertising service it self. It could be that once one service disables access for piracy websites, the operators of those websites switch to the best competitor. Second, smaller firms may be less likely to join the initiative, perhaps because they are not aware of European-level policy, perhaps because their cost structure does not allow sufficient leeway to participate in policy meetings. The fact that there are many more small firms than large firms in the sample may affect the average effect. It is possible that small firms disproportionately rely on industry associationss for representation in European-level policy interactions. However, the *Guiding Principles* make it clear that industry associations do not

Table 1: Baseline results: Evaluation of the "Guiding Principles" agreement

	(1)	(2)	(3)
After × non-EU Ad	-0.009		-0.019
	(0.025)		(0.054)
After × EU Ad	-0.043		-0.192**
	(0.072)		(0.081)
After × non-EU Ad × Popularity: low		-0.026	
		(0.061)	
After × non-EU Ad × Popularity: medium		0.012	
		(0.027)	
After × non-EU Ad × Popularity: high		-0.030	
		(0.038)	
After × EU Ad × Popularity: low		-0.028	
		(0.053)	
After × EU Ad × Popularity: medium		0.055	
		(0.091)	
After × EU Ad × Popularity: high		-0.549***	
		(0.082)	
Weighted by Popularity	No	No	Yes
Observations	72571	72571	72571
$\overline{R^2}$	0.900	0.905	0.973

Dependent variable: Log number of piracy websites served per third-party service.
After indicates the period after October 21, 2016, *EU Ad* indicates advertising services that use a TLD of an EU-country,
non-EU Ad indicates advertising services that do not use a TLD of an EU-country.
The sample spans 3 months prior to the agreement and 12 months after the agreement.
All models include a group-specific linear time trend, third-party service and time period fixed effects.
Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

have any responsibility to transfer the obligations to their members. Even if small firms are more likely to be members, they might still be less likely to comply because of the aforementioned reasons. In what follows, we test for these explanations by allowing for heterogeneity by the number of websites the third-party service is connected to. We interpret this definition of *popularity* as both a measure of differentiation and a measure of firm size (and therefore cost structure).

In column (2) we explicitly allow for heterogeneity by estimating a flexible specification that distinguishes between three classes of third-party services based on their overall popularity, i.e. the number of websites that use the tracker in a given snapshot of the *HTTPArchive*.[26] We account for low, medium

---
[26]Note that we cannot distinguish between more size classes because cell sizes would quickly get too small to draw

and high *non-EU Ad* and *EU Ad* services. A third-party service of *Popularity: low* is requested by $\leq$100 websites, *Popularity: medium* is defined as >100 but $\leq$2000 websites, and third-party services of *Popularity: high* are requested by >2000 websites. As can be seen in the descriptive statistics in table A.1, in the treatment group of most interest, *EU Ad*, the distribution of these size classes is about 28% (low), 57% (medium) and 15% (high) with respect to the total number of EU-based advertising services. The market for third-party services, however, is highly concentrated and a small number of third-party services are requested by a large number of websites. When we look at the market share in the piracy market, advertising services that we classify of low popularity type are requested by 1.5% of piracy websites, advertising services in the medium popularity category are requested by 6.9% of piracy websites and the most popular advertising services are requested by 36.0% of piracy websites. Looking at the subsample of EU-based advertising services, the respective shares are 1.2%, 2.8%, and 6.9%.

The estimation results show that the number of piracy websites connected to advertising services only changes in the case of the most popular EU-based advertising services. The estimate in column (2) suggest a decrease of 42%.[27] Also in this specification, we do not find evidence for indirect effects of the self-regulation, i.e. the number of piracy websites that interact with non-EU advertising services does not change significantly, independently of the popularity of those services.

In the average effect in column (1), we have treated all third-party services equally, and didn't take into consideration that the probability that we observe a given service in our sample varies substantially because of the concentrated market structure. The specification doesn't reflect the idea that some third-party services are more likely to be picked by piracy websites. This can be true for several reasons. First, it may be more costly to search for, locate, and establish trust with an advertising service that is less popular. Second, in the presence of network effects, there can be important incentives to join an advertising intermediary that offers access to a larger number of advertisers. The specification in column (1) additionally doesn't reflect the idea that for a public policy effort to be considered successful, it is not necessarily required that the average agent responds accordingly, but perhaps much more that the most important agents do so. While the heterogeneity analysis in column (2) provides some remedy to the apparent fallacy of division, we provide a perhaps more

---

meaningful inference. Furthermore, a linear specification would force a restrictive functional form on the data that wouldn't allow to illustrate the key result as prominently.

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

elegant way in column (3). Here we report results from a regression that assigns probability weights to individual observations based on the overall popularity (i.e. number of connected websites) of third-party services. In this specification, more popular, and therefore perhaps more important, third-party services are given greater weights in the estimated average effect. In line with our heterogeneity results, the estimated effect of the self-regulation on EU-based advertising services in this specification is a decrease of 17.5%. Because of its convenience and parsimony in terms of degrees freedom, and because we are much more interested in a qualitative rather than quantitative interpretation of the results, we make the weighted regression specification our preferred specification for the additional analyses below.

Throughout all specifications in table 1, we don't see any evidence that the indirectly affected firms, i.e. non-EU advertising services, reacted to the self-regulation effort. In addition to the firm size effect which we have established, this suggests that there is also important heterogeneity in terms of firm type. Only the directly affected firms, and out of those only the largest ones, comply to the self-regulation in way that we can pick up in our dataset and using our identification strategy.

4.2 Dynamic perspective and test of the parallel trend assumption

One might be concerned about a potential threat to exogeneity of the quasi experiment. 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 advertising-related and not advertising-related services 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 advertising services to non-advertising services, snapshot by snapshot:

$$
\begin{aligned}
Log(NumberWebsites_{it}^k + 1) = \alpha + \sum_t \beta_0^t w_t + \sum_t \beta_1^t \left(w_t \times \textit{EU-Ad}_i \right) \\
+ \sum_t \beta_2^t \left(w_t \times \textit{Non-EU-Ad}_i \right) + \mu_i + \varepsilon_{it}.
\end{aligned} \tag{2}
$$

If advertising third-party services adapt their behavior in expectation of the agreement, we would expect significant β^τ coefficients before the *Guiding Principles*. Estimation results, alongside with 90% confidence bands are depicted in figure 3. The estimates broadly suggest that the number of piracy

17

Figure 3: Group differences over time

EU-Ad vs. Control   *Non-EU-Ad vs. Control*
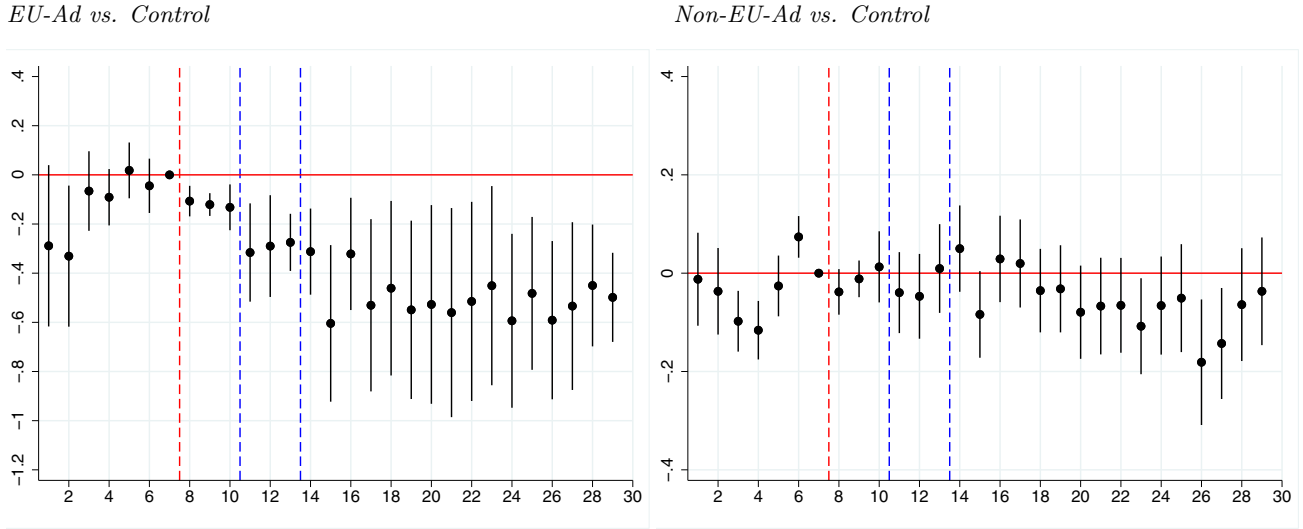
Differences between the treatment group (EU-Ad/non-EU-Ad) and the control group (non-Ad services).
Vertical axis: *OLS coefficients.* **Horizontal axis:** *Data snapshots 2016–2018.*
OLS estimates of the $\beta_1^t w_t$ (left panel) and $\beta_2^t w_t$ (right panel) coefficients obtained from a regression of:
$Log(NumberPiracyWebsites_{it} + 1) = \alpha + \sum_t \beta_0^t w_t + \sum_t \beta_1^t (w_t \times \text{EU-Ad}_i) + \sum_t \beta_2^t (w_t \times \text{Non-EU-Ad}_i) + \mu_i + \varepsilon_{it},$
weighted by tracker popularity (see discussion in section 4.1.) Bars indicate 90% confidence bands.
Dashed vertical lines indicate the Guiding Principles *of October 21, 2016 (red line), three months after the agreement (first blue line), and six months after the agreement (second blue line). Note that we don't observe the same number of data snapshots for each pre- and post-period (see sections 3.2.*

websites served by advertising services did not change differently than the number of piracy websites served by non-advertising services prior to the agreement (i.e. prior to the first red dashed vertical line). Note, however, that the identifying assumption of the difference-in-differences estimator seems to be better met regarding EU-based advertising services (the 'direct' treatment group) compared to non-EU-based advertising services (the 'indirect' treatment group). By definition, this analysis is of course the same from the perspective of an evaluation of the *Memorandum of Understanding* signed on June 25, 2018, because we would still want to compare affected and unaffected firms after this event to the period before *any* intervention has happened.

Further, this exercise also shows that there is no clear change in the number of piracy websites that make requests to non-EU-based advertising services after the experiment, neither in the short run nor in the long run. The results are very different when we look at the number of piracy websites that make requests to EU-based advertising services. Here we see an immediate significant decrease relative to the control group, which becomes stronger over time and eventually stagnates.

We now turn to investigate whether the results above are specific to the time window of 3 months

18

Table 2: Different time windows

	(1)	(2)	(3)
After × non-EU Ad	0.011	0.009	-0.019
	(0.033)	(0.042)	(0.054)
After × EU Ad	-0.138**	-0.079	-0.192**
	(0.062)	(0.050)	(0.081)
Weighted by Popularity	Yes	Yes	Yes
Observations	24815	32467	72571
$\overline{R^2}$	0.979	0.976	0.973
Sample	± 3 months	−3/+6 months	−3/+12 months

Dependent variable: Log number of piracy websites served per third-party service.
After indicates the period after October 21, 2016, *EU Ad* indicates advertising services that use a TLD of an EU-country, *non-EU Ad* indicates advertising services that do not use a TLD of an EU-country. *Popularity* is measured as the number of websites that use the third-party service. All models include a group-specific linear time trend, third-party service and time period fixed effects. Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

prior and 12 months after the *Guiding Principles* agreement. To aid comparison, we repeat the results reported in column (3) of table 1 in column (3) of table 2. In column (1), we restrict the sample to ± 3 months and arrive at very similar results. We find a significant decrease in the number of piracy websites connected to EU-based advertising services in the order of 12.9%. Compared to the estimate in column (3), this is 4.6 percentage points smaller. However, this difference is not statistically significant in the sense that 90% confidence bands overlap. In column (2), we restrict the sample to a 3 month pre-period and a 6 months post-period. In this setting, we do not find a significant effect for neither non-EU based nor EU-based advertising services. Looking at the left-hand side panel of figure 3, which plots snapshot-by-snapshot differences between the control group and EU-based advertising services, suggests that this is due to less precise estimates of the effect in the months 4–6 after the agreement (the period between blue dashed lines). The general notion of a reduction in the number piracy websites, in a weighted regression framework, therefore also remains when we restrict our attention to this sample.

4.3 Placebo and counterfactual 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

Table 3: Placebo: Baseline effect and different timing

	(1)	(2)	(3)
After × non-EU Ad	0.024	-0.032*	-0.034
	(0.015)	(0.017)	(0.024)
After × EU Ad	0.064	0.061	0.067
	(0.077)	(0.081)	(0.072)
Weighted by Popularity	Yes	Yes	Yes
Observations	24815	32467	72571
$\overline{R^2}$	0.991	0.988	0.985
Sample	± 3 months	−3/+6 months	−3/+12 months

Dependent variable: Log number of placebo websites served per third-party service.
After indicates the period after October 21, 2016, *EU Ad* indicates advertising services that use a TLD of an EU-country, *non-EU Ad* indicates advertising services that do not use a TLD of an EU-country. *Popularity* is measured as the number of websites that use the third-party service. All models include a group-specific linear time trend, third-party service and time period fixed effects. Standard errors in parentheses, clustered on the third-party service level.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

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 3, we restrict the sample to ± 3 months. In column (2), we define the post-period to include 6 months, and in column (3), the sample includes the 12 month post-period that we also used in our preferred baseline specification. We find that there is not much evidence for a general trend that can explain the results in table 1. We do not find a significant difference in the number of placebo websites that make requests to EU-based advertising services, independent on the time window. Except for a marginally significant negative coefficient in the 6 month window, we also do not find that non-EU-based advertising services change their connections to placebo websites. Comparing the coefficients in column (2), we conclude that EU-based and non-EU-based advertising services do not differ significantly in the number of connected placebo websites. In summary, the results in table 3 imply 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 run the same specification as in column (5) of

Table 4: Counterfactual: One year before the "Guiding Principles" agreement

	(1)	(2)
After × non-EU Ad	0.025	0.001
	(0.026)	(0.024)
After × EU Ad	-0.014	-0.074
	(0.106)	(0.059)
Weighted by Popularity	Yes	Yes
Observations	13174	18785
$\overline{R^2}$	0.997	0.996
Sample	± 3 months	−3/+6 months

Dependent variable: Log number of piracy websites served per third-party service.

After indicates the period after October 21, 2015, *EU Ad* indicates advertising services that use a TLD of an EU-country, *non-EU Ad* indicates advertising services that do not use a TLD of an EU-country. *Popularity* is measured as the number of websites that use the third-party service.

Note that the sample is based on 7 snapshots of historical data based on HTTP request as listed in the HTML versions of the archived pages on *CommonCrawl*. All models include a group-specific linear time trend, third-party service and time period fixed effects.

Standard errors in parentheses, clustered on the third-party service level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

table 1. The coefficient of $After \times EU\,Ad$ in column (1) of Table 4 is small and not significantly different from zero. The same holds true in column (2), where we use a longer sample of 6 months post-counterfactual-experiment. Note that we cannot estimate a version of this specification with a window of 12 months, because this would include the factual experiment. We also don't see evidence for an increase in the number of piracy websites connected to non-EU-based advertising services.

4.4   Robustness checks

In this section we briefly discuss some further robustness checks, mostly concerning potential measurement error and data quality issues, before we turn to an attempt of an evaluation of the *Memorandum of Understanding* and a discussion of the broader implications of our results.

4.4.1   Heterogeneity in the control group

As discussed above, the control group of non-advertising services consists of a broad variety of different types (see the distribution in figure A.1). This may affect the comparison with the treatment group and therefore the estimates of the difference-in-differences coefficients. In columns (1)–(4) of table A.2, we exclude a category of non-advertising third-party services each and re-estimate the baseline

model. The results remain very similar, which leads us to conclude that there is not much evidence that the composition of the control group has substantial effects on our results. It further clearly suggests that non-advertising services as a whole did not react to the self-regulatory effort in either a direct nor indirect way.

4.4.2 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.3 show that the estimated effect remains qualitatively similar to those of our baseline results in column (1). Some specifications yield significantly larger (more negative) estimates of the

22

effect size. Note that the number of observations varies across these different measures because of missing data. Further, we estimate a specification where we use information on the signatories of the *Memorandum of Understanding* as a measure of the unknown signatories of the *Guiding Principles*. As discussed above, a number of reasons suggest that there might not be perfect overlap between these two measures. First, it is possible that some firms were part of the first agreement, but did not sign the MoU later. Second, some firms that were not part of the first agreement, might have not signed the MoU later. Third, some of the signatories are interest groups and industry bodies that have no legal power to force their members to comply with the MoU. We again find a negative coefficient of *After* × *Signatory Ad* in column (5), but the estimate is not significantly different from zero. Given the fact that there is not a extremely strong correlation between the two measures (see also the discussion in section 5.1), it is not surprising that using the MoU signatory list as a proxy for the signatories of *Guiding Principles* introduces some noise to the estimation.

4.4.3 Intensive and extensive margins

Our analysis has focused on a specific measure of the intensive margin, i.e. the number of piracy websites that send requests to a given third-party service. In table A.4, we report results of an alternative measure. In column (2), we look at the share of requests a third-party service receives from piracy websites relative to the requests it receives in total. We find the same pattern as in our baseline specification. Relative to the mean of the dependent variable (0.00384), a coefficient of −0.001 translates into a reduction of 28.7%. In column (3) we investigate whether the agreement to the *Guiding Principles* has had an effect on the extensive margin. The dependent variable here indicates whether a given third-party service receives any requests from piracy websites at all.[28] The estimated coefficients suggest that there is no effect on the extensive margin. Hence, we do not find evidence for a strategic reaction, in the sense that some advertising services that didn't serve the piracy market before, start to do so in response to the self-regulation.

4.5 Extension: The "Memorandum of Understanding"

In section 4.2, we have shown that the effects of the *Guiding Principles* agreement persisted for at least one year. We now go on to study whether this trend continues after the final version of the

---

[28] We observe third-party services that do not receive requests from piracy websites by construction of our sample: about 42.2% of the services in the sample only receive requests from our placebo websites.

Table 5: Evaluation of the "Memorandum of Understanding"

	Log(NumberPiracy)	Log(NumberPlacebo)	NumberWebsites=0	
	(1)	(2)	(3)	(4)
AfterGP × non-EU Ad	-0.070	-0.021	0.011	0.013
	(0.060)	(0.028)	(0.008)	(0.008)
AfterGP × EU Ad	-0.258**	0.120	-0.014	-0.013
	(0.116)	(0.109)	(0.042)	(0.041)
AfterMoU × non-Signatory Ad	-0.168***	-0.114***	0.052***	
	(0.060)	(0.043)	(0.017)	
AfterMoU × Signatory Ad	-0.272***	-0.121**	0.002	
	(0.101)	(0.060)	(0.011)	
AfterMoU × non-EU Ad				0.029**
				(0.012)
AfterMoU × EU Ad				0.012
				(0.038)
Weighted by Popularity	Yes	Yes	Yes	Yes
Observations	116596	116596	453698	453698
$\overline{R^2}$	0.966	0.982	0.636	0.636

Dependent variable: Log number of piracy websites served per third-party service.

AfterGP indicates the period after October 21, 2016, *AfterMoU* indicates the period after June 25, 2018.

EU Ad indicates advertising services that use a TLD of an EU-country, *non-EU Ad* indicates advertising services that do not use a TLD of an EU-country, *Signatory Ad* indicates advertising services that have signed the MoU, *non-Signatory Ad* indicates advertising services that did not sign the MoU. *Popularity* is measured as the number of websites that use the third-party service.

The sample covers the period June 2016 to August 2018. All models include a group-specific linear time trend, third-party service and time period fixed effects.

Standard errors in parentheses, clustered on the third-party service level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

self-regulatory agreement (*Memorandum of Understanding*), which was signed on June 25, 2018.

The list of signatories is publicly available, hence we can have a more direct measure of whether a third-party service is directly affected by the *MoU*. The identification strategy and econometric model remains very similar to equation (1). In the specification in column (1) of table 5 we add a second set of difference-in-differences coefficients that capture a difference after the MoU, and is estimated on an extended sample that covers the period June 2016 to October 2018.

Looking at the coefficients of *AfterGP × non-EU Ad* and *AfterGP × EU Ad* coefficients first, we again see the familiar patterns of the baseline results. There is no evidence that the number of piracy

websites changes for indirectly affected advertising services, but we find a significant decrease in the number of piracy websites that are connected to directly affected advertising services. The estimated effect corresponds to a decrease of 22.7% as a response to the agreement to the *Guiding Principles*. The point estimate is similar when we consider the estimate of the effect of the MoU on signatory advertising services (23.8%). Strikingly, we find a negative and significant coefficient of *AfterMoU ×* *non-Signatory Ad*. Before we can conclude that this implies that the MoU has had spillover effects on firms that didn't participate, let's again consider a placebo exercise.

Much like in section 4.3 above, a convincing placebo exercise would yield small and insignificant coefficient estimates for all difference-in-differences coefficients. In column (2), we find that this is not the case, neither concerning *AfterMoU × non-Signatory Ad*, nor *AfterMoU × Signatory Ad*. The fact that we find a significant coefficient regarding both piracy and placebo websites suggests that there is an unobserved general trend that coincides with the timing of the MoU. A potential cause of such a trend could be the *General Data Protection Regulation*, which came into force on May 25, 2018 and demands higher standards of consumer data privacy, protection and portability.[29] As a result of this regulation, in which non-compliance is sanctioned with substantial monetary penalties, it is possible that some advertising services exited the (European) market. Column (3) provides some support for this conjecture. Here we estimate a linear probability model, trying to understand whether the probability of serving zero websites in a given snapshot changes for *Signatory Ad* and *non-Signatory Ad* services. We find a positive and significant coefficient of *AfterMoU × non-Signatory Ad*, which implies a 2.9 percentage points higher likelihood of those firms exiting the market. In column (4) we show that this result holds when we look at advertising services with EU-based TLDs. Putting these insights together, we are less confident that we can identify the causal effect of the MoU. If one is willing to believe that the difference between the coefficient in the piracy model (column 1) and the placebo model (column 2) can be interpreted as a slightly less problematic estimate of the causal effect, these results would suggest that piracy websites reduced their requests to non-signatory advertising services by 5.3% and requests to signatory advertising services by 14.0%.

---

[29]The notion of "data protection by design and by default" in the General Data Protection Regulation implies that firms are required to (pseudo-)anonymize personal data where appropriate and that explicit informed consent is needed to process personal data. The regulation applies to firms within the EU as well as to firms outside the EU if they interact with individuals in the EU. More details can be found at https://ec.europa.eu/info/law/law-topic/data-protection_en.

5   Discussion

After we have established the effects of the agreement(s) to self-regulate, we now turn to a discussion of the potential mechanisms that can explain the results. We also speculate about the self-regulation's success in reaching its overarching goal: a decrease in online piracy.

5.1   The popularity effect

Our results suggest that service popularity 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. In addition, 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. Finally, larger companies may find it easier to bear the costs of implementing steps to reduce the connections to piracy websites, and to continue to monitor whether banned/unlicensed websites continue to sign up for advertising intermediary services.

There is no publicly available information on which firms participated in the stakeholder meetings or have signed the agreement of *Guiding Principles*. The publicly available list of signatories of the *Memorandum of Understanding*, however, provides some interesting insight. Table 6 provides a descriptive view of the determinants of being a signatory of the MoU. We find that more popular services are more likely to sign the MoU. Further, with *advertising* as the omitted category, this exercise suggests that all other types of services are less likely to sign the MoU. We also find a higher likelihood for members of an industry association. This correlation is perhaps not surprising given that even the official press release of the *Guiding Principles* agreement 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*.[30] Some firms might 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. We do not find that services that have received a positive number of requests from piracy websites before the *Guiding Principles* agreement have a

---

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

Table 6: Determinants of signing the Memorandum of Understanding

	(1)	
Hosting/CDN	-0.160***	(0.009)
Unknown	-0.143***	(0.009)
Tracking	-0.145***	(0.013)
Interactive	-0.101***	(0.039)
IndustryAssociation	0.598***	(0.013)
Log(NumberWebsites)	0.005***	(0.001)
PiracyConnection	0.006**	(0.003)
EU TLD	0.026***	(0.007)
Constant	0.128***	(0.010)
Observations	7189	
$\overline{R^2}$	0.408	

Dependent variable: Signatory of the *Memorandum of Understanding* (0/1).
Advertising is the omitted category, top-level domain fixed effects included.
Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

lower likelihood of signing the MoU – in fact, we find a positive correlation. Finally, the results also show a correlation between using a EU-based top-level domain and being a signatory of the MoU.

5.2 Overall effectiveness

Finally, we will speculate whether the intervention was successful in reaching its overarching goal: a reduction in online piracy. A way to measure this, is to measure the demand for piracy websites.

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, 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 7. We do not find much evidence in favor of this mechanism. Across different time horizons in columns (1)–(3), the estimates of $After \times EU\,Ad$ are not significantly different from zero. In the long time horizon (which

Table 7: Pageviews of piracy websites and the Guiding Principles agreement

	(1)	(2)	(3)
After × non-EU Ad	0.016	0.039	0.072*
	(0.025)	(0.025)	(0.040)
After × EU Ad	0.048	0.036	0.090
	(0.046)	(0.044)	(0.077)
Weighted by Popularity	Yes	Yes	Yes
Observations	4922	6537	13984
$\overline{R^2}$	0.957	0.933	0.909
Sample	± 3 months	−3/+6 months	−3/+12 months

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 tracker 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.
* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

corresponds to our preferred baseline specification in table 1), we even find a positive and borderline significant coefficient of $After × non − EU Ad$. We conclude that this exercise does not provide very convincing evidence that the self-regulation has substantially reduced the demand for online piracy.

5.3 Parallels to other markets

As indicated in section 2.2, online advertisers and online advertising service intermediaries are just one of the four areas in which the European Commission seeks to establish soft law measures to reduce the incentives for commercial IP infringement. The basic mechanics that we highlight in this paper – coming from important heterogeneity in terms of firm type and firm size, as well as that we don't find much evidence for spillover effects beyond the European Union – are informative for the design of future initiatives for self-regulation. Although there are differences in the market structure of online advertising intermediaries, online payment and logistics and transportation intermediaries, the overall effectiveness of such measures might be improved by establishing greater incentives to join for small firms, specialized firms, or firms outside of the European Union. The latter seems especially relevant in the context of online advertising, given its highly concentrated market structure and the fact that European advertising services are by far less popular than non-European advertising services.

Further, self-regulation is not uncommon in the advertising industry and its underlying economics

appear to be more general.[31] Therefore, our results also provide implications for different settings where the availability of online advertising creates incentives to produce content that has negative externalities. To illustrate this idea, consider an example that seems to resemble remarkable similarities with the results we present in this paper.

Much like in the market for unlicensed software and media content, the supply of false and misleading information ("fake news") is – anecdotally – mainly funded by advertising.[32] The observation that a number of news articles about US politics, which are based on objectively false information, were widely shared on social media created a public debate on whether fake news could have biased voting decisions in the 2016 presidential election.[33] 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.[34] 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).[35] However, anecdotally it seems that the strategic incentives to self-regulate can vary across firms. Considerably 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".[36]

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

---

[31]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.

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

[33]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.

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

[35]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.

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

and legitimate websites. We compare the dynamics before and after the advertising industry, under pressure from the European Commission, has agreed to a self-regulation that aims 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 popularity, 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. We find that the number of piracy websites that interact with the most popular EU-based advertising services decreases by 42%, results from a regression that assigns more weight to more popular third-party services yields an estimate of an average effect of -18%. We show that non-advertising services, out of which a substantial fraction is in the business of collecting, analyzing (and indirectly 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 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 visits to piracy websites, and conclude that there is no evidence of a downward trend on the demand-side.

While our study mainly contributes to the literature of 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

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. (2018). "Catch Me If You Can: Effectiveness and Consequences of Online Copyright Enforcement." *Information Systems Research*, *29*(3), 656–678.

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. (2018). "The Impact of Consumer Multi-homing on Advertising Markets and Media Competition." *Management Science*, *64*(4), 1574–1590.

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

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.

Estrada-Jiménez, J., Parra-Arnau, J., Rodríguez-Hoyos, A., and Forné, J. (2017). "Online Advertising: Analysis of Privacy Threats and Protection Approaches." *Computer Communications*, *100*, 32–51.

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.

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. (2018). "On the Ubiquity of Web Tracking: Insights from a Billion-Page Web Crawl." *Journal of Web Science, 4*(4), 53–66.

Shiller, B., Waldfogel, J., and Ryan, J. (2018). "The effect of ad blocking on website traffic and quality." *The RAND Journal of Economics, 49*(1), 43–63.

Sun, M., and Zhu, F. (2013). "Ad revenue and content commercialization: Evidence from blogs." *Management Science, 59*(10), 2314–2331.

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.

Figure A.1: Distribution of third-party services

Relative frequency of observations   *Number of websites*
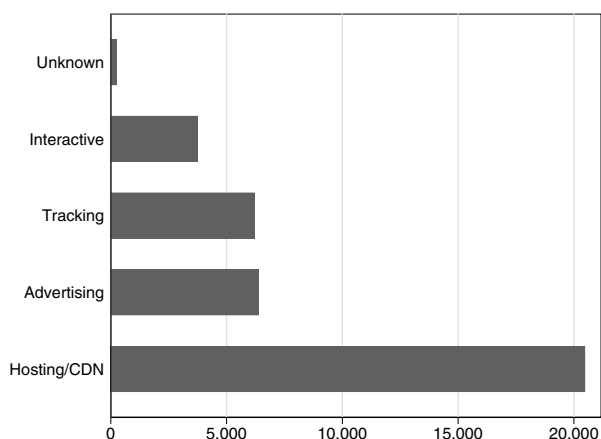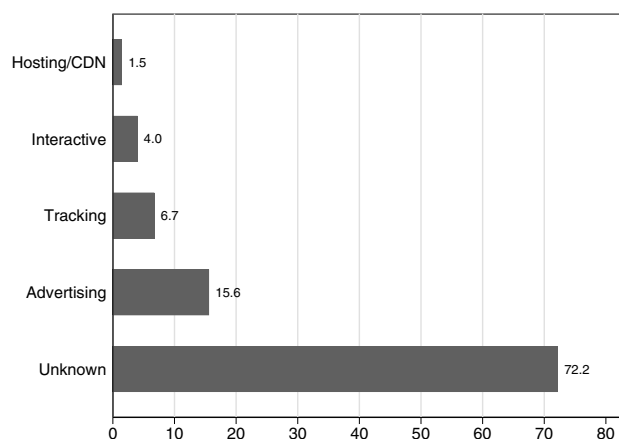
Left panel*: Percent of observations in the sample.*
Right panel*: Average number of websites served per category in the sample.*
The sample spans 3 months prior to the agreement and 12 months after the agreement.

Table A.1: Descriptive statistics

| | non-EU Ad (N=487), EU Ad (N=43), Control (N=6,659) | | | | | | | | | | | |
| | non-EU Ad Before | | non-EU Ad After | | EU Ad Before | | EU Ad After | | Control Before | | Control After | |
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
NumberPiracyWebsites	5.374	12.773	5.029	12.329	2.087	4.945	1.838	3.726	1.466	8.898	1.318	7.856
log(NumberPiracyWebsites)	0.991	1.161	0.975	1.123	0.595	0.846	0.578	0.839	0.450	0.657	0.411	0.645
NumberPlaceboWebsites	14.449	41.686	13.942	40.303	3.339	6.057	3.582	5.512	2.815	20.301	2.910	19.685
log(NumberPlaceboWebsites)	1.701	1.230	1.734	1.188	0.958	0.876	1.104	0.819	0.672	0.712	0.721	0.713
NumberWebsites	6558.099	21127.572	6892.006	22744.852	1228.061	2834.439	1340.320	2747.901	1220.276	11493.288	1338.826	12126.803
log(NumberWebsites)	7.095	1.856	7.266	1.785	5.545	1.680	5.759	1.627	3.950	2.357	4.190	2.367
Popularity: low	0.075	0.264	0.065	0.246	0.283	0.451	0.281	0.449	0.604	0.489	0.563	0.496
Popularity: medium	0.540	0.498	0.518	0.500	0.584	0.493	0.562	0.496	0.333	0.471	0.365	0.481
Popularity: high	0.385	0.487	0.417	0.493	0.133	0.340	0.157	0.364	0.063	0.243	0.073	0.259

The sample spans 3 months prior to the agreement and 12 months after the agreement.

34

Table A.2: Different types of non-advertising services

	(1)	(2)	(3)	(4)
After × non-EU Ad	-0.032	-0.023	-0.000	-0.020
	(0.054)	(0.059)	(0.053)	(0.056)
After × EU Ad	-0.144*	-0.143*	-0.192**	-0.194**
	(0.080)	(0.075)	(0.082)	(0.082)
Weighted by Popularity	Yes	Yes	Yes	Yes
Observations	20167	67692	69676	71504
$\overline{R^2}$	0.974	0.967	0.973	0.969
Excluded	Unknown	Tracking	Interactive	Hosting

Dependent variable: Log number of piracy websites served per tracker.

After indicates the period after October 21, 2016, *EU Ad* indicates advertising services that use a TLD of an EU-country,

non-EU Ad indicates advertising services that do not use a TLD of an EU-country.

Popularity indicates the number of websites that use the third-party service.

All models include a group-specific linear time trend, tracker and time period fixed effects.

Standard errors in parentheses, clustered on the tracker level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

35

Table A.3: Different measure of treatment groups

	(1)	(2)	(3)	(4)	(5)
After × non-EU Ad	-0.019				
	(0.054)				
After × EU Ad	-0.192**				
	(0.081)				
After × non-EU(Reg) Ad		-0.016			
		(0.055)			
After × EU(Reg) Ad		-0.560***			
		(0.109)			
After × non-EU(IP) Ad			-0.037		
			(0.068)		
After × EU(IP) Ad			-0.319***		
			(0.107)		
After × non-EU(LinkedIn) Ad				-0.020	
				(0.054)	
After × EU(LinkedIn) Ad				-0.318***	
				(0.106)	
After × non-Signatory Ad					0.014
					(0.045)
After × Signatory Ad					-0.056
					(0.084)
Weighted by Popularity	Yes	Yes	Yes	Yes	Yes
Observations	72571	33195	28441	38315	72571
$\overline{R^2}$	0.973	0.973	0.974	0.974	0.973
Sample					

Dependent variable: Log number of piracy websites served per tracker.

After indicates the period after October 21, 2016.

Signatory Ad-Service indicates advertising trackers that (indirectly) signed the final MoU,

non-Signatory Ad-Service indicates advertising trackers that did not (indirectly) sign the final MoU.

Popularity indicates the number of websites that use the tracker.

All models include a group-specific linear time trend, tracker and time period fixed effects.

Standard errors in parentheses, clustered on the third-party service level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Table A.4: Different measure of dependent variable

	(1) Log(NumberPiracy)	(2) Share Piracy	(3) Piracy (0/1)
After × non-EU Ad	-0.019	0.000	-0.001
	(0.054)	(0.000)	(0.009)
After × EU Ad	-0.192**	-0.001**	0.001
	(0.081)	(0.000)	(0.020)
Weighted by Popularity	Yes	Yes	Yes
Observations	72571	72571	72571
$\overline{R^2}$	0.973	0.909	0.775

Dependent variables: Log number of piracy websites served per tracker (column 1),

Share of piracy websites within all websites served per tracker (column 2),

Indicator whether a tracker serves no piracy websites at all (column 3).

After indicates the period after October 21, 2016, *EU* indicates a TLD of a EU-country.

Popularity indicates the number of websites that use the tracker.

All models include a group-specific linear time trend, tracker and time period fixed effects.

Standard errors in parentheses, clustered on the tracker level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$

Table A.5: Different sample, not excluding questionable data points

	(1)	(2)	(3)
After × non-EU Ad	0.011	-0.034	-0.031
	(0.033)	(0.048)	(0.052)
After × EU Ad	-0.138**	-0.138**	-0.132
	(0.062)	(0.062)	(0.082)
Weighted by Popularity	Yes	Yes	Yes
Observations	24815	35755	75859
$\overline{R^2}$	0.979	0.968	0.969
Sample	± 3 months	−3/+6 months	−3/+12 months

Dependent variable: Log number of piracy websites served per tracker.

After indicates the period after October 21, 2016. *EU Ad-Service* indicates advertising trackers that use a TLD of an EU-country, *non-EU Ad-Service* indicates advertising trackers that do not use a TLD of an EU-country. *Popularity* indicates the number of websites that use the tracker. All models include a group-specific linear time trend, tracker and time period fixed effects. Standard errors in parentheses, clustered on the tracker level.

* $p < 0.10$, ** $p < 0.05$ *** $p < 0.01$