

What Do News Aggregators Do?

Evidence from Google News in Spain and Germany^{*}

Joan Calzada[†]

Ricard Gil[‡]

December 2018

Abstract

The impact of aggregators on news outlets is ambiguous. In particular, the existing theoretical literature highlights that although aggregators create a market expansion effect when they bring visitors to news outlets, they also generate a substitution effect if some visitors switch from the news outlets to the aggregators. Using the shutdown of the Spanish edition of Google News in December of 2014 and difference-in-differences methodology, this paper empirically examines the relevance of these two effects. We show the shutdown of Google News in Spain decreased the number of daily visits to Spanish news outlets between 8% and 14%, and that this effect was larger in outlets with less overall daily visits and a lower share of international visitors. We also find evidence suggesting that the shutdown decreased online advertisement revenues and advertising intensity at news outlets. We then analyze the effect of the opt-in policy adopted by the German edition of Google News in October of 2014. Although such policy did not significantly affect the daily visits of all outlets that opted out, it reduced by 8% the number of visits of the outlets controlled by the publisher Axel Springer. Our results demonstrate the existence of a net market-expansion effect through which news aggregators increase consumers' awareness of news outlets' contents, thereby increasing their number of visits.

^{*} We thank Shane Greenstein, Avi Goldfarb, Hal Varian, Susan Athey, Omar Nayeem, Mike Bailey, Lisa George, Paul Seabright, Lesley Chiou, Gerard Llobet, Jorge López, Alvar Teixidor, and audiences at the Winter NBER Meeting on the Economics of Digitization, Eighth Annual Conference on Internet Commerce and Innovation at the Searle Center of Northwestern University, New Economic School in Moscow, Johns Hopkins Carey Business School, American University, the Federal Communications Commission, Universitat de Barcelona, Universitat Oberta de Catalunya, the 10th Conference on "The Economics of Intellectual Property, Software and the Internet" at Toulouse, and the FSR Communications & Media, CMPF and ENTraNCE Annual Scientific Seminar on the Economics, Law and Policy of Communications and Media in Florence, Italy, and the First Catalan Economic Society Meetings in Barcelona, Spain, for their valuable comments and suggestions. The usual disclaimer applies.

[†] Universitat de Barcelona. Av. Diagonal 690, Barcelona 08034, Spain; calzada@ub.edu.

[‡] Smith School of Business, Queen's University. Goodes Hall, 143 Union St W, Kingston, ON, CANADA; ricard.gil@queensu.ca.

What Do News Aggregators Do?

Evidence from Google News in Spain and Germany.

Abstract

The impact of aggregators on news outlets is ambiguous. In particular, the existing theoretical literature highlights that although aggregators create a market expansion effect when they bring visitors to news outlets, they also generate a substitution effect if some visitors switch from the news outlets to the aggregators. Using the shutdown of the Spanish edition of Google News in December of 2014 and difference-in-differences methodology, this paper empirically examines the relevance of these two effects. We show the shutdown of Google News in Spain decreased the number of daily visits to Spanish news outlets between 8% and 14%, and that this effect was larger in outlets with less overall daily visits and a lower share of international visitors. We also find evidence suggesting that the shutdown decreased online advertisement revenues and advertising intensity at news outlets. We then analyze the effect of the opt-in policy adopted by the German edition of Google News in October of 2014. Although such policy did not significantly affect the daily visits of all outlets that opted out, it reduced by 8% the number of visits of the outlets controlled by the publisher Axel Springer. Our results demonstrate the existence of a net market-expansion effect through which news aggregators increase consumers' awareness of news outlets' contents, thereby increasing their number of visits.

1. Introduction

Online platforms and aggregators have drastically changed how consumers and businesses gain access to information and interact with each other. Because consumers now use aggregators and search engines to find all sorts of goods and services such as flights, accommodation, or insurance, firms have adapted their online distribution channels to the growing presence of aggregators to remain competitive. The academic community has also noticed the importance of this shift in consumption patterns and business practices; consequently, there is a growing number of papers in economics and management aiming to understand the role of aggregators in online markets.

A type of aggregator that has rapidly grown in importance is the one simplifying the search of news stories such as Google News, Yahoo! News, Bing News, or Summify. News aggregators offer links to news stories published by news outlets, which are usually complemented with excerpts and images. They allow consumers to save considerable time and effort in finding news. In spite of the growing appeal of these websites to internet users, traditional news outlets around the world have been reticent to their introduction because of their potential effects on the consumers' browsing behavior and, consequently, on their advertisement revenues. Even though news outlets can opt-out of aggregators by using software that blocks the links to their content, most publishers want to be indexed while receiving some economic compensation for the use of their content. This new scenario has generated multiple frictions between Google and publishers in Europe, and has led to changes in intellectual property laws in several countries.¹ This paper examines two recent events of disputes of Google News in Spain and Germany that prove useful to understand the role and impact of news aggregators on the news market in different ways. Whereas Google News completely shut down its Spanish edition, in Germany a group of publishers voluntarily decided to reduce their presence in the German edition of Google News.

The theoretical literature has identified an important trade-off in the various effects of news aggregators on news outlets' audiences (Dellarocas et al., 2013; George and Hogendorn, 2012; Calzada and Ordoñez, 2015; Jeon and Nasr, 2016; Jeon, 2018). On the one hand, aggregators create a “market-expansion effect” because they allow consumers to discover news outlets with low brand awareness that otherwise they would not know. Moreover, aggregators may induce consumers to read news stories of higher quality from several outlets and dedicate more time overall to reading. The number of page views that news outlets obtain from links in news aggregators depends on several factors such as the font size of their headlines in the aggregator, the number of words in the excerpts accompanying their links, or the use of images. These factors affect both the interest of consumers in clicking through to the original link and reading the full news story (Dellarocas et al., 2016). On the other hand, the presence of aggregators may lead some consumers to quit browsing the front page of news outlets. This generates a “substitution effect” that depends

¹ Google launched Google News in 2002 and today is present in more than 70 countries in 35 different languages. Additionally, it also has many regional editions that are adapted to the specific needs of the audiences.

on the outlets' brand loyalty. Therefore, determining whether the indirect page views generated by the market-expansion effect compensate for the loss of direct visits by the substitution effect, and their overall effect on revenues, is a relevant empirical question. Moreover, from a policy perspective, it is also important to understand how aggregators change consumers' engagement habits, and which news outlets are more likely to benefit from news aggregators.

This paper sheds light on these questions by first analyzing the impact of the shutdown of Google News in Spain in December of 2014. At the beginning of 2014, a reform of the Spanish intellectual property law established that firms posting links and excerpts of news stories have to pay a compulsory link fee (Google tax) to the original publishers. Consequently, on December 16, 2014, Google News decided to shut down its Spanish edition, arguing that under the new regulation, this service would not be profitable. We exploit this quasi-natural experiment to examine the effect of news aggregators on the number of visits received by news outlets and on the consumers' engagement metrics.

We then complement this analysis by considering a change in the linking policy of Google News adopted in Germany around the same time. In 2013, the German Parliament introduced a change in the copyright law that allowed news aggregators to link for free the news stories of news outlets if using excerpts of less than 7 words. The use of longer excerpts or images would require the payment of a negotiated fee to the news outlets. After a number of disputes with some German outlets, Google News changed from an "opt-out" to an "opt-in" policy. Under the new opt-in rule, those German publishers that want to be indexed by Google News must give up receiving any compensation from the aggregator. After this change in Google News' linking policy, on October 23, 2014, an association of publishers named VG Media decided to protect their contents by not opting in. Soon after, they reversed their decision allegedly due to drastic traffic losses, and accepted Google's opt-in conditions.

Our study draws from a rich data set obtained from SimilarWeb containing information for 151 newspapers in Spain, Germany, France, and Italy. This data set includes information about each domain's daily number of visits, and several engagement metrics per day (daily average page views per visit, visit duration, and bounce rate). We complement these data with information on advertisement revenues from a sample of Spanish outlets from Arce Media.

Assuming as plausibly exogenous the source of variation provided by Google's shutdown in Spain and its opt-in policy in Germany and using French and Italian news outlets as control groups, we apply a model of difference-in-differences to assess the role of this aggregator in the news market. Our results show that after the Google News' shutdown in Spain, Spanish news outlets experienced a reduction in the number of daily visits between 8% and 14%, with a growing impact during the first six weeks. Our findings confirm previous analysis in the empirical literature showing that news aggregators have a net positive impact in outlets' traffic (Chiou and Tucker, 2017).

We also find that reductions in daily visits after the shutdown are larger in lower-ranked outlets and outlets with a lower percentage of international visitors. These results are consistent with the fact that news aggregators benefit the most those outlets that have low brand awareness and a large brand loyalty. The shutdown affected the most lower-ranked outlets and less international outlets because they are most dependent on the flow of casual consumers generated by aggregators, and because aggregators do not compete for their relatively small base of loyal domestic visitors. Along the same lines, we also show the impact of the shutdown depends on the outlets' specialization. Specifically, we find the shutdown affected sports and regional outlets the most, had a lower effect on national outlets, and did not significantly affect business outlets. As far as engagement metrics are concerned, we find the shutdown of Google News reduced the average duration and the number of pages per visit while increasing bouncing rates. These findings suggest the shutdown significantly changed the composition of the consumers that visit news outlets.

Our analysis of the effect of the shutdown on advertisement revenues and advertising intensity at the domain level merits a careful description. On the one hand, our empirical analysis compares advertising revenues and advertising intensity at the site-advertiser-date level across news and non-news Spanish domains before and after the shutdown of Google News in Spain. We show that both advertisement revenues and advertisement intensity in news outlets decreased after the shutdown. This impact is weaker on front pages than content pages, which receive most traffic from news aggregators.

The analysis of the German case shows that Google News' opt-in policy generated a negligible effect on the visits to the news outlets that decided to stay out. Only when we restrict the analysis to the news outlets controlled by the publisher Axel Springer, we find a 8% reduction in daily visits. When Google adopted the opt-in policy, Axel Springer and other publishers promoted the opt-out option, but only the outlets in the VG Media consortium followed. Consequently, the loss of visits to these outlets shows the existence of a "competitive effect" related to their shorter excerpts. Eventually, the reduction of visits and the salient loss of visibility in the national ranking for some of its flagship newspapers motivated Axel Springer to finally accept Google's opt-in conditions.

A few papers have analyzed the effects of aggregators in the news market.² Athey and Mobius (2012) examine the impact that introducing local news headlines and links in Google News can have on consumers' browsing behavior. In 2009, the French edition of Google News enabled a local news feature that allowed those users that entered their zip code to obtain news from local outlets. Using a data set of consumer browsing behavior, the authors compare users who adopted the localization feature with a sample of control users that exhibited similar consumption patterns in the past. They find the addition of local news content led users to rely more on Google when initiating a browsing session. Moreover, after the introduction of this feature, direct navigation to local outlets increased by 5%

² See Goldfarb and Tucker (2017) for a review of the literature analyzing the effects of copyright laws in digital markets.

(bypassing Google News altogether), and clicks on local outlets from the Google News page increased by 13%. In a related paper, George and Hogendorn (2013) use a major redesign in the US edition of Google News on June 30, 2010. Similar to the previous case, the redesign placed a permanent strip of geo-targeted local news headlines onto the Google News front page. Using a sample of news visits by US households before and after the introduction of the geo-targeted links, the authors find local news visits increased by less than 1% and the likelihood of a local news visit increased between 4% and 6% from a low baseline for heavy Google News users. Interestingly, the results show no evidence of substitution away from direct outlet visits. Adding geo-targeted links increased the number of different local outlets visited per day, but not the number of unique sites visited per month. Finally, Chiou and Tucker (2017) analyze the impact of a contract dispute between Google News and The Associated Press (AP) when Google News removed all AP news articles from December 23, 2009 until sometime in February 2010. Using weekly data on the top 150 sites users navigated immediately after visiting Google News or Yahoo! News, they find Google News users were less likely to visit other news websites after visiting Google News following the removal of AP content, relative to Yahoo! News users, who did not experience such a content change.

Our paper adds to this literature by providing evidence of the change in daily visits per news outlet after the Spanish edition of Google News stopped operations. This natural experiment allows us to compare the Spanish outlets (treatment group) with French or Italian news outlets (control groups) for which no changes of news aggregators occurred. In addition, we use consumers' engagement metrics to examine the navigation habits of outlets' visitors.³ We complement the analysis of this event with the examination of the opt-in policy of Google in Germany. This case allows us to measure whether the impact of news aggregators on news outlets depends on how the news aggregator exhibits information. In the German case, we compare the daily visits of those German outlets that decided to opt-out (treatment group), with French or Italian news outlets, and with those German news outlets that decided to opt-in (control group). The singularity of the two events described, and the granularity of the data analyzed, allows us to test the predictions from the theoretical literature and our section on mechanisms. Moreover, our analysis of the effect of the shutdown on the news outlets' advertisement sales is important to understand the economic relevance of news aggregators for the survival of online news sites.

To the best of our knowledge, the closest paper to ours is Athey, Mobius and Pal (2017), which independently study the impact of the Google News shutdown in Spain, using individual-level browsing data, and comparing Google News users and non-Google News users before and after the shutdown. The results of both papers are in fact complementary in that their results are consistent with ours (both papers find an overall 10% reduction in the number of daily visits to news outlets), even though the data and methodology differ

³ Aggregators may reduce the number of loyal users and increase the number of casual consumers. This substitution between user types can reduce advertising revenues at some news outlets (Athey, Calvano, and Gans, 2011).

across papers. On the one hand, our data is at the domain level including information from all visits from desktop users, and their data is at the individual level for only users of Microsoft products, accounting for half of PC news browsing activity. On the other hand, our methodology compares a sample of Spanish news outlets to a control group of French and Italian news outlets, and theirs use individual-level browsing data to compare a designated group of Google News users to non-Google News users in Spain. Additionally, our paper provides two extra pieces of evidence that offer a wider view of the impact of news aggregators in the media market. First, we examine the impact of the shutdown on Spanish outlets' advertising revenue and intensity. Thus, we are able to assess the financial impact of news aggregators on news outlets beyond the impact on site traffic. Second, we also study the effects of Google's opt-in policy in Germany, which allows analyzing the "competitive effect" that appears when news aggregators give a differentiated treatment to the links of news outlets.

Finally, our paper is also related to other recent research that has examined the effects of news aggregators and indexing beyond the Google News case. Roos et al. (2015) investigate the influence of excerpts on consumers' decisions to consume news. They show that observing just one excerpt reduces consumers' uncertainty about their match with the excerpted site's content by about 33%. They conclude that excerpting benefits the linked site by increasing the share of traffic originating at the linking site, and benefits the linking site by making it more popular at the start of consumers' browsing sessions. The paper also finds that excerpting increases news consumption, leading consumers to browse more frequently and visit a wider range of sites. Relatedly, Cage et al. (2015) examine 84 general information media outlets in France (including newspapers, television channels, radio stations, and news agencies), and track every article these sites offered online in 2013, with the help of a plagiarism-detection algorithm that quantifies the copy rate between an article and all the articles previously published about the event. They find that half of online information production is copy-and-paste. They also explain that those outlets that produce more content receive more visits, but the rapid spillover of information occurring in the last few years has reduced the incentives of newspapers to produce original news stories. Chesnes et al. (2017) analyze the impact of the Google's ban from sponsored search listings in 2010 on pharmacies not certified by the National Association of Boards of Pharmacy (NABP). Using difference-in-differences and synthetic control methodology, they show that the ban increased the search costs for non-NABP-certified websites, but some consumers overcame this increase in search costs by switching from sponsored to organic links for other-certified websites. Finally, Sismeiro and Mahmood (2018) study the impact of an exogenous Facebook outage on the number of visits to a large online news website in a major Western European country. Their results are consistent with the literature in that they also find evidence of a net market-expansion effect of online social networks on news outlets visits. Their paper points out that, despite these similarities, social networks differ from news aggregators in two ways. First, while aggregators rely on algorithms, online social network rely on friends to filter and recommend content. Second, because the scope of activities performed at online social networks is far richer, complementarities between

news stories and recommending parties may be an important driver on clicking behavior and reading patterns.

The rest of the paper is structured as follows. Section 2 describes the main institutional details of the Spanish and German cases. Section 3 outlines the mechanisms behind the impact of news aggregators on news outlets, describes our data, and presents our methodology. Section 4 analyzes the shutdown of Google News in Spain. Section 5 assesses the impact of the opt-in policy adopted by Google News in Germany. Finally, section 6 concludes.

2. Institutional Details: Google's Disputes with European Publishers

Since the release of Google News in 2002, news publishers around the world have fought against the free indexation of their content while advocating for receiving some economic compensation from Google. Even though this situation has generated several legal disputes, some European governments have recently considered creating a link fee (Google tax) that would force news aggregators to compensate the linked outlets.⁴ This section first reviews the earlier history of the relationship between Google and the European news publishers, and later describes the creation of a link fee in Spain and Germany, which motivate our empirical analysis.⁵

Belgium was among the first countries to regulate the activities of news aggregators. In 2006, Copiepresse (representing French-language and German-language Belgian publishers) sued Google News over alleged copyright infringement. Consequently, in 2006 and 2007, two sentences forbid Google News to link the contents of Belgian publishers without their consent.⁶ In 2011, the Belgian Appeals Court ratified these decisions and established that the mere linking of newspaper websites should be considered infringement. Soon after this resolution, the Belgian publishers asked to be linked back to Google News, and on December 12, 2012, Google agreed to index Copiepresse newspapers under the condition of no future legal action for copyright infringement. The agreement also established that the two parts would collaborate on several business initiatives to promote both the publishers' and Google's services.⁷ Thierry Geerts, Google Belgium's managing director, clearly announced how Google aimed to address similar disputes in other

⁴ Similarly, in September of 2016 the European Union announced its intention to reform the copyright legislation, which might entail the creation of the so-called neighboring rights to protect the contents of press publishers. See news release here, http://europa.eu/rapid/press-release_IP-16-3010_ca.htm.

⁵ In 2012 the National Association of Newspapers in Brazil persuaded its 154 members to ban Google News to use its contents, arguing that Google was refusing to pay for the links and was driving traffic away from their websites. In the following years some news outlets allowed back links from the aggregator. See news release here, <http://www.bbc.com/news/world-latin-america-20018221>.

⁶ Today, Google offers news outlets the option to opt out of Google News if they feel harmed by the links. See, for example, the agreement between Google and the Italian anti-trust authorities in 2011. See news release here, http://www.nytimes.com/2011/01/18/technology/18iht-google18.html?_r=2.

⁷ See news release here, <http://www.theverge.com/2012/12/13/3764692/google-copyright-lawsuit-settlement-belgium>.

countries⁸: “Instead of continuing to argue over legal interpretations, we have agreed on the need to set aside past grievances in favor of collaboration. This is the same message we would like to send to other publishers around the world - it is much more beneficial for us to work together than to fight.”

Similar to the Belgian case, several publishers in France lobbied the French government in 2012 to create a link fee. Google reacted to this initiative by threatening to close its French edition if this measure were approved. By February 1, 2013, French President François Hollande and Google Executive Chairman Eric Schmidt reached an agreement such that French publishers agreed to forego the establishment of a link fee and Google agreed to create a €60 million Digital Publishing Innovation Fund that would support transformative digital publishing initiatives for French readers. Google also offered to help French publishers increase their online revenues using its advertising technology, which allow for better targeting of consumers.⁹

More generally, an overall revision of Google News' disputes in Europe shows Google's strategy has been to lobby against the establishment of a link fee while investing numerous resources to gain the publishers' support. A clear example of such strategy is the launch by Google on April 2015 of the Digital News Initiative (DNI), collaborating with Les Echos (France), FAZ and Die Zeit (Germany), the Financial Times and the Guardian (UK), NRC Group (Netherlands), La Stampa (Italy), and El Pais and Grupo Godo (Spain). DNI will dedicate €150 million to projects that support innovation in digital news journalism over the next three years, and will invest in training and development resources for journalists and newsrooms across Europe.¹⁰

2.1. The Shutdown of Google News in Spain

The main dispute between Google News and the European publishers took place in Spain. On January 1, 2014, the Spanish Parliament passed a reform of the Law of Intellectual Property (LPI).¹¹ The new law established that online outlets posting links and excerpts of news articles originated elsewhere must pay a link fee (canon) to the original publishers. The creation of the link fee was initially promoted by the publishers association AEDE (Asociación de Editores de Diarios Españoles), which lobbied the government to force news aggregators to compensate them for the use of their content.¹²

⁸ See news release here, <http://googlepolicyeurope.blogspot.de/2012/12/partnering-with-belgian-news-publishers.html>.

⁹ See news release here, <https://googleblog.blogspot.fr/2013/02/google-creates-60m-digital-publishing.html>.

¹⁰ See news release here, <http://googleespana.blogspot.com.es/2015/04/google-y-editores-de-medios-de-europa.html>.

¹¹ See news release here, <https://www.boe.es/boe/dias/2014/11/05/pdfs/BOE-A-2014-11404.pdf>.

¹² The passing of this regulation was not free of controversy. Whereas some of the biggest Spanish publishers argued in favor of it, others, such as AEEPP (Asociación Española de Editoriales de Publicaciones Periódicas) and Coalición Pro-Internet, opposed it. The Spanish regulator CNMC (Comisión Nacional de los Mercados y

A unique feature of the Spanish regulation is that publishers cannot refuse to receive a fee from news aggregators; in fact, the link fee must be collected by a private entity called CEDRO, which will distribute back the revenues to the news outlets. This strategy tries to prevent publishers from giving away their right to receive compensation, and to enforce coordination among publishers. Note that if the fees were voluntary, some publishers could negotiate exclusivity agreements with Google and put their rivals at a competitive disadvantage.

Although the implementation of the law involved a lot of uncertainty, on December 11, 2014, Richard Gingras, world responsible of Google News, unexpectedly announced that on December 16, Google News would shut down its Spanish edition.¹³ Google justified this action by claiming that the new regulation made the service unprofitable because Google News had no direct source of revenues (the firm does not show any advertising on this site).¹⁴ Google's decision was shortly followed by other, yet smaller, Spanish news aggregators, such as Planeta Ludico, NiagaRank, Multifriki, InfoAliment, and Beeinfo. Others tried to modify their content to avoid the effects of the law (Planet Ubuntu, Astrofisica, and Fisica).

The shutdown of Google News had an important and immediate impact on the Spanish news market. Some early reports estimated a reduction in the daily visits of the largest newspapers of more than 8%, and even bigger for smaller newspapers (NERA, 2015 and 2017). As a result, the publishers in AEDE and other associations have urged the government to negotiate a solution with Google.¹⁵ Some large publishers in AEDE have even announced they would renounce any compensation payment for sharing content with news aggregators. In spite of this backlash, the solution to this case may be delayed until the European Commission approves its new copyright legislation, which could modify the regulatory framework to protect publishers in the European Union.

2.2. The Opt-in policy of Google News in Germany

The second case we examine in this paper is a dispute between Google News and the German news publishers. On March 1, 2013, the German Parliament passed an addendum to the copyright law that granted publishers the right to charge search engines and other online aggregators for reproducing their content, but the law also allowed the free use of text in links and brief excerpts. This addendum meant publishers can prohibit aggregators from using their news articles beyond headlines and short excerpts, and they can charge aggregators a link fee if the aggregators make a larger use of their contents. The main

la Competencia) also advocated for the modification of several aspects of the new regulation. See CNMC (2014) and Llobet (2015).

¹³ See news release here, <http://googleespana.blogspot.com.es/2014/12/novedades-acerca-de-google-noticias-en.html>.

¹⁴ See news release here, <https://support.google.com/news/answer/6140047?hl=es>.

¹⁵ See news release here, <http://www.aede.es/wp-content/uploads/2015/02/AEDEPrensa-CierreGoogleNewsDic14.pdf>.

differences in this regulation concerning the policy adopted in Spain a few months later are that (1) link fees have to be negotiated between the parties, and (2) it does not affect brief excerpts.

In June 2014, VG Media,¹⁶ a consortium of more than 200 publishers, including Axel Springer, sued Google and other news aggregators for displaying excerpts and preview images along with the links to their news articles. VG Media alleged that aggregators were using their content without their consent, and that they should receive compensation according to the new law.¹⁷ Google refused to pay the publishers, and instead modified its linking policy. On October 2, 2014, the German edition of Google News announced the change from an opt-out to an opt-in system. This change implied that those German publishers that want to be indexed by Google News must explicitly grant permission and renounce any type of compensation.¹⁸

After this change in Google's policy, publishers, and TV and radio stations associated with VG Media decided not to opt in. A leading publisher in this group was Axel Springer, which asked VG Media not to issue free licenses for its websites (welt.de, computerbild.de, sportbild.de, and autobild.de). Other publishers that followed the same course of action were Burda (bunte.de), Funke, Madsack, and M. DuMont Schaubergas. Phillip Justus, Managing Director of Google Germany, answered that Google "will not show in the future snippets and thumbnails of the publishers members of VG Media."¹⁹

On October 23, 2014, Google News and other German news aggregators stopped showing large excerpts, video, and images from the publishers that did not opt in to avoid paying them a link fee. This change allegedly significantly reduced the number of daily visits VG Media news sites received from Google and overall. Mathias Döpfner, Axel Springer Chief Executive, estimated that the downgrading of search notices resulted in a loss of nearly 40% in traffic volume, and that the traffic from Google News was down by almost 80%. Moreover, welt.de dropped below its competitors in the IVM and AGOF rankings, and computerbild.de lost its Top 10 rank of all AGOF offerings in Germany.²⁰ Shortly after, on

¹⁶ In December 2015, VG Media had 332 members. See a list here, <https://www.vg-media.de/de/alle-wahrnehmungsberechtigte.html>.

¹⁷ VG Media accused Google of "extortion" and "discrimination" against its members, saying the law passed the previous year guaranteed news providers a "reasonable compensation for the use of their content." See news release here, <http://phys.org/news/2014-10-google-news-snippets-germany-legal.html>.

¹⁸ Google continues to use an opt-out policy outside Germany, implying that publishers must take several actions to prevent indexing of their contents. For example, they must use a robot.txt file or metadata to let Google's crawlers know what to ignore.

¹⁹ See news release here, <https://germany.googleblog.com/2014/10/news-zu-news-bei-google.html>.

²⁰ Axel Springer SE detailed that the financial damage resulting from lost marketing revenues could be in the seven-figure range per brand over the year as a whole. See the post here, http://www.axelspringer.de/en/presse/Axel-Springer-concludes-its-data-documentation-Major-losses-resulting-from-downgraded-search-notices-on-Google_22070687.html.

November 5, 2014, Axel Springer and other VG Media publishers decided to opt in and gave Google a license to add excerpts to their search results free.²¹

3. Mechanisms, Data, and Empirical Methodology

3.1. Mechanisms

When analyzing the role of news aggregators, we borrow our framework from the existing theoretical literature. We assume as departure point that consumers visit a limited number of news outlets, which typically differ on the topics they report. Under standard assumptions, news aggregators may cause up to three different effects in the news market. First, news aggregators may increase the number of daily visits for indexed news outlets. According to the literature, this market expansion effect occurs because aggregators offer contents of a higher quality and variety than traditional news outlets, which induce consumers to read more news stories (Jeon and Nasr, 2016) and to increase the total amount of reading time instead of doing other leisure activities (Dellarocas et al. 2013). In this context, the different treatment aggregators give to news outlets can alter their number of daily visits, and have an additional effect on market competition (Dellarocas et al. 2016). Second, aggregators may create a substitution effect that reduce the number of direct visits to indexed news outlets. A substitution effect occurs when the consumers that visit news aggregators may not visit the news outlets' front pages or click through to their content pages (Dellarocas et al. 2016). Third and last, aggregators may also have an information screening or matching role. Consumers can observe the quality of news stories by reading the excerpts in the aggregator, and can click on the links only if the perceived quality is high (Huang, 2017). Consequently, consumers can stay longer reading the news outlets pages because of better news match. This matching role may not increase the number of visits or traffic per se but it may modify the composition of the audience of news outlets.

We want to empirically evaluate the role of news aggregators in the media market by analyzing how the previous mechanisms affect the visits to news outlets. Because our data is aggregated at the domain-day level, we cannot investigate the impact of the shutdown of Google News in Spain on consumer behavior, as Athey, Mobius and Pal (2017) did with consumer-level data. By contrast, we can estimate the net impact of the shutdown on the number of visits to different types of news outlets that were exposed to different intensities of the market expansion and substitution effects. We also investigate the relevance of the matching role by estimating the impact of the shutdown on different engagement metrics such as pages per visit, visit duration and bounce rates at the domain-day level.

²¹ Mathias Doepfner used this case to illustrate Google's ability to modify competition in the media market: "As sad as it is, at least now we know precisely how enormous the consequences of discrimination are, how Google's market power really plays out, and how Google punishes those who exercise the right to protect content." See news release here, <http://www.ibtimes.com/news-media-no-match-google-axel-springers-about-face-1720494>.

To examine the mechanisms driving differences in the net impact of the market-expansion and substitution effects, we consider the role of brand awareness and brand loyalty in determining news outlets' audiences. On the one hand, brand awareness reflects how much and whether consumers know about an outlet, and therefore whether they spontaneously browse over its news content. News seekers usually consider a small set of news outlets when searching for news stories. As a result, aggregators may allow consumers to become aware of additional outlets and to read more news stories (Jeon, 2018). Following this logic, we predict the lower the brand awareness of an outlet, the larger the market expansion effect the aggregator generates for that particular outlet.

On the other hand, brand loyalty reflects how often a consumer repeatedly visits the same news outlet, regardless of the content offered by other news outlets, due to their ideological preferences or editorial affinities. Consumers with brand loyalty experience a reduction in utility when they have a strong preference for a news outlet and switch to an aggregator. Under this framework, we predict that the lower the news outlets' brand loyalty, the smaller the consumers' preference mismatch and the larger the substitution effect the aggregator creates for that particular outlet.

Under the umbrella of brand awareness and brand loyalty, we can determine how aggregators may affect news outlet's daily traffic. In a first scenario, consider a news outlet that has both a large brand awareness and brand loyalty. For this type of outlet, news aggregator will have little effect in the number of daily visits because most readers already know about their existence (no large market expansion effect). Note also that with multi-homing loyal consumers will continue visiting these outlets (no large substitution effect). This profile would match that of predominant national outlets, and business news outlets, which cannot be replaced by aggregators due to the readers' ideological preferences or because they are niche outlets with strong loyal audience. In our empirical model, we will identify the former type as the most popular outlets (measured by outlet rank and percentage of foreign visitors), and the latter as those outlets specialized on a particular niche with poor substitutes.

In a second hypothetical scenario, imagine an outlet with low brand awareness and a large brand loyalty. In such a case, the outlet will benefit from the presence of news aggregators because it will receive visits from casual consumers that otherwise would not be aware of its presence (market expansion effect). In contrast, the outlet would not lose visitors due to its consumers' loyalty (no substitution effect). Potentially, this profile could match the case of regional and sports outlets, which have a loyal reader base and gain page views with the extra visitors provided by the aggregator. While regional outlets may attract visitors from other regions, sports outlets may receive visits from multi-homing visitors loyal to other sports outlets. There is horizontal differentiation among regional and sports outlets in that they specialize in regions and sports teams based in different cities. This means that while

they have a strong loyal audience, they also receive many visits through news aggregators from consumers based in other regions.²²

A final and third hypothetical scenario is one where an outlet has low brand loyalty, combined with a low or a high brand awareness. Given the low brand loyalty of the outlet, some consumers will switch to the aggregator if available (large substitution effect). However, according to the theoretical literature, these outlets may benefit from news aggregators if they offer differentiated contents and their brand awareness is sufficiently small (large market expansion effect). This hypothetical outlet may match the profile of low rank traditional national outlets or new online outlets with an unclear editorial policy, which can benefit from aggregators because their articles complement the contents offered by generalist national outlets. In addition, note that branding will play an important role for these outlets when news aggregators disappear because causal news seekers, who used to rely on aggregators, will end up visiting those outlets with a higher profile.

Note that in the different scenarios examined above the key determinant for the effect of news aggregators is the relative importance of brand awareness and brand loyalty, together with the outlets' generalist and niche specialization. According to our mechanism analysis, we can derive the following three testable implications about the effects of the shutdown of Google News on news outlets daily visits. First, news outlets with large brand awareness and large brand loyalty (high-rank outlets, international reputed outlets, predominant national outlets and business outlets) will be less affected by the shutdown. Second, news outlets with low brand awareness and high brand loyalty (regional and sports outlets) would be relatively more affected by the shutdown. Third, news outlets with low brand awareness and low brand loyalty (low-rank national outlets) will be more affected by the shutdown. Finally, we also test whether news aggregators generate an information screening effect and induce readers to visit more page views and read for longer periods. If this is the case, even if aggregators had no effect on the number of daily visits we should observe differences in the visits characteristics such pages viewed per visit, visit duration, and bounce rates after the shutdown of Google News in Spain.

We should be clear about the fact that when considering the impact of the shutdown on the different types of news outlets identified above, we assume that the algorithm used by Google News to select and rank news stories does not discriminate across outlets for reasons other than content quality and appeal.²³

²² Regional outlets have high loyalty and high brand awareness among regional readers, but on the national stage there is likely less brand awareness so when one considers brand awareness as an average across the entire population, this measure would have a low value. A similar logic applies to sports outlets (in Spain at least) because sports outlets tend to support teams in a particular geographical region. This means that among people living in that region or supporters of team in that region, brand awareness may be high, but at a national level is low. Consumers that multi-home only browse one sports outlet (even if they know about the existence of others), but when access a news aggregator they may read news stories of news outlets that they will usually not consider.

²³ Joel Sommerland explains quite intuitively how Google's algorithm selects the contents linked to the Google News webpage: *"The algorithm reviews content automatically, looking for indicators of quality, assessing a story's placement based on the number of user clicks it is attracting, the popular consensus on the*

In the next subsections, we present our data and methodology that we use to test our predictions. Our empirical analysis mainly aims to estimate the relative importance of the market expansion and substitution effects using differences on the impact of the shutdown of Google News in Spain across news outlets.

3.2. Data

In this study, we use data at the domain-day level from SimilarWeb, a web measurement company providing traffic data and user-engagement statistics. This firm collects data on browsing behavior from rich and diversified panels of consumers in several countries. The data we use come exclusively from desktop users. The information covers the period from June 1, 2014, to May 31, 2015, which includes the two events analyzed in the paper. Google News' shutdown on December 16, 2014 affected Spanish news outlets. Therefore, our data cover roughly half a year before and after Google News' shutdown in Spain. Google's remove of excerpts and images from October 23, 2014, to November 5, 2014, affected the German news outlets belonging to the VG Media consortium.

To explore the impact of the Google News shutdown in Spain on news outlets, we chose French and Italian news outlets as control groups after considering other countries such as Germany and Portugal. While discarding Germany as a control group was a no-brainer because of its own Google News "related turbulence" in October 2014, Portugal is very different from Spain in terms of its population size, broadband usage, internet usage, and other demographics and consumption characteristics of internet consumers.²⁴ Table 1 shows that Spain's internet penetration and household broadband access rates are far higher than Portugal. Despite the notable differences in population size, Spain, France and Italy have similar percentages of households with access to broadband telecommunications services, and Internet users for the whole population and per age bracket. Spain appears ranked in the middle according to the use of e-banking and emails among countries in Table 1, and only exceeding others in terms of its use of social networks. This overall comparison makes French and Italian news outlets the most adequate control groups for our exercise out of the available Western European countries. Regarding our analysis of the German market, we start using French and Italian news outlets as comparison groups to German news outlets for similar reasons. Because only VG Media sites in Germany opted

trustworthiness of its publisher, the relevance of the story to the reader's current geographical location and the freshness (i.e. publication date and time) of the story in question. Google News is therefore more likely to rank British news sites highly when the story concerns a fire in London than reports on the same incident from much-admired publishers from further afield like The New York Times or Washington Post. The recurrence of specific keywords across publications and the level of public interest indicated by user searches guide the algorithm in its creation and organisation of specific subjects into clusters." See Joel Sommerland, The Independent, June, 18 2018. <https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-news-headlines-stories-ranking-algorithm-editors-publishers-journalism-a8404811.html>

²⁴ Another reason is the fact that news outlets in Portugal receive much less visits than news outlets in Spain, France and Italy due to the smaller population size in Portugal. SimilarWeb data are less precise when the number of daily visits is below 5000 and so the number of Portuguese news outlets with reliable daily information is far smaller than in other countries such as Italy and France.

out from Google News, our analysis takes us in the end to compare VG Media and Axel Springer outlets (a publisher within VG Media) to other German news outlets in our sample.

We have identified and selected news outlets in our sample according to their national rankings published by Alexa (www.alexa.com) and SimilarWeb (www.similarweb.com). We picked top rated news outlets excluding webpages from TV and radio stations, and other potential news aggregators such as MSN or Yahoo. In order to classify news outlets, we searched for verbal descriptions in several sources such as Alexa, SimilarWeb and Wikipedia.²⁵ Overall, we aimed to have a well-balanced sample of news outlets classified in different categories such as their specialization (national, regional, business, or sports), their rank at the national level and their internationalization level (their percentage of domestic versus foreign visitors).

In the end, our data set contains information for 151 domains, including 50 news outlets from Spain, 32 from Germany, 29 from France, and 40 from Italy. Table 2 offers a complete listing of all domains. We also have information about the Spanish, German, French, Italian and Portuguese editions of Google News, about Yahoo! News in Spain (es.noticias.yahoo.com), and about two additional Spanish news aggregators (meneame.net and kiosko.net). All domains are classified according to different criteria. First, we categorize them according to their specialization. They can be National, Regional, Business, or Sports. Second, we divide domains according to their national rank. Specifically, we distinguish between the Top 50% and the Bottom 50% of domains of our sample. Third, we classify domains according to the number of visits they receive from other countries. Top International outlets are those that receive more than 25% of the visits from abroad (we set the threshold for Italian outlets at 11% because they have far fewer international visits). Top International 50% and Bottom International 50% separate the outlets of the sample into two groups according to whether their share of international visits is above or below the median in our sample. Finally, in the case of Germany, we also consider whether the domains belong to the VG Media consortium and whether Axel Springer (completely or partly) owns them. Table 2 reports the list of domains analyzed with their specialization, and whether they are in the Top or Bottom 50% of their country in our sample.

The main variable of our analysis is the domain's Daily Visits. This variable is defined as the daily entries to a web domain from a different web domain or from the beginning of an empty browsing session, and expires after 30 minutes of inactivity. We also consider several engagement metrics. Visit Duration is the session length, which is the time that elapses between the first and the last page visit, on the analyzed domain. Note that according to this definition, the visit duration is equal to zero when the visitor only visits one page within the domain. On the other hand, during the duration of the visit, all the activities such as clicking on articles and images are considered page views. Pages per Visit is the daily page views divided by the daily visits of the domain. Finally, Bounce Rate is a

²⁵ See for example, https://en.wikipedia.org/wiki/List_of_newspapers_in_France.

variable that measures the percentage of daily single-page sessions out of all daily sessions for the domain. This variable measures how often a consumer reaches a web page and then leaves without navigating to any other page. In such instances, the visitor stays in the domain for a very short period of time.

Tables 3A and 3B shows the summary statistics for all the variables obtained from SimilarWeb for our Spanish and German analyses, respectively. See from Table 3A that the average site in our Spanish, Italian and French data receives 245,000 daily visits. On average, users see 3.7 pages during their visits that lasts around 10 minutes (622,8 seconds). The average site in our German analysis data receives 302,000 daily visits, and each visitor sees 3.8 pages per visit in 7 minutes (428,4 seconds).

Tables 4A and 4B separate the data by country, and also before and after each of the Google turbulences under study. Note that French, Italian and German domains have similar characteristics to the Spanish domains in our sample. This observation validates the use French and Italian domains as control groups for Spanish news outlets in our empirical analysis. In fact, pages per visit and bounce rates in Spanish, French and Italian news outlets are very similar. The proportions of national, regional, business and sports in each country sample also resemble each other. Outlets in all four countries have similar percentages of domestic visits and, if anything, our variable definition means German and Italian outlets are less likely to appear among the top international outlets in our full sample. This fact may be because more people in the world speak and read French and Spanish than they speak and read German and Italian.

Note that Spanish domains have on average fewer daily visits than French, Italian and German domains in our sample. This can reflect that our sample contains a larger number of Spanish sites with lower rankings. See also in Table 4A that all countries experience a decrease in daily visits after the shutdown of Google News in Spain, and that the decrease in Spain and Italy are proportionally larger than the decrease in France. Interestingly, Table 4B shows that visits in German outlets increased during the opt-out period but decreased elsewhere in France and Italy. Therefore, leveraging differences in domain type (national, regional, business or sports, top or bottom 50%, top or bottom of international visits) is important to understand mechanisms underlying the impact of the shutdown of Google News on news outlets' traffic.

3.3. Empirical Methodology

We use a difference-in-differences methodology to investigate both the impact of the shutdown of Google News in Spain and the opt-in policy of Google News in Germany. Next, we present the methodology used in our analysis of the Spanish case. Our first specification examines the total impact of the Google News shutdown on the news outlets' daily visits and consumers' engagement, therefore capturing the joint net effect of the elimination of the market-expansion and the substitution effects, and the information screening effect. Although we expect the shutdown to eliminate the outlets' visits of news

seekers from the aggregator, some of these consumers could directly visit the outlets and modify the consumers' overall navigating behavior. We identify the net effect of the shutdown on the domains with the following model:

$$\ln[y_{ijt}] = \theta SPAIN_i * Shutdown_t + \mu_{j0}t_{jt} + \mu_{j1}t_{jt}^2 + \gamma_i + \delta_t + u_{ijt},$$

where $\ln[y_{ijt}]$ is the log of the outcome and dependent variable (e.g., daily visits to site i in country j in day t), $SPAIN_i$ is a dummy variable that takes the value of 1 if site i belongs to Spain, and 0 otherwise. $Shutdown_t$ is another dummy variable that takes the value of 1 if day t is after December 16, 2014, and 0 otherwise. We also use country-site specific fixed effects γ_i and date fixed effects δ_t , and introduce country-specific time trends t_{jt} and t_{jt}^2 to control for differences in long-term trends between Spain and the control group. Finally, we assume the error term u_{ijt} to be iid and normally distributed as usual. The main objective of our analysis is to identify the impact of the shutdown in the daily visits, but we also analyze other engagement visitor metrics such as average pages viewed per visit per day and site, average duration of visits per day and site, and average bounce rate per day and site. We want to investigate whether visitors of news aggregators and news outlets' direct visitors have different navigation habits, and thus we test whether news outlets experienced differences in the number of pages per visit and the duration of the average visit after the shutdown.

Because we use country-site specific fixed effects and date fixed effects in all specifications, the dummies $SPAIN_i$ and $Shutdown_t$ are not separately identified. Our parameter of interest is θ (the diff-in-diff parameter), and it captures the net effect of Google News' shutdown on Spanish news outlets. Therefore, the treatment group is all Spanish outlets in our sample, and the treatment period is the days after December 16, 2014. We run two separate exercises with different control groups, French and Italian news outlets, during the same period. While the demographics and habits of internet users in France and Spain are similar, the Charlie Hebdo terrorist attack shortly after the shutdown of Google News presents identification concerns. For this reason, the results using Italian news outlets as control group strengthen the validity and robustness of our overall findings.

Our second specification divides the impact of the treatment by week from the first week to seventh week and beyond after the shutdown of the Spanish edition of Google News. It is as follows:

$$\ln[y_{ijt}] = \sum_{k=1}^7 \theta_k SPAIN_i * k^{th}AfterShutdown_t + \mu_{j0}t_{jt} + \mu_{j1}t_{jt}^2 + \gamma_i + \delta_t + u_{ijt},$$

where θ_k is a parameter that captures the net effect of the shutdown on Spanish newspapers in a day t within the k th week after the shutdown of Google News. All other parameters, variables and fixed effects remain the same from the explanation above. If anything, in our specifications using French news outlets as the control group, we introduce a dummy for the fourth week after the shutdown in French newspapers. We do so to control for the unanticipated increase in the number of visits French news outlets received due to the Charlie Hebdo attack in Paris. These events took place in the fourth week after the shutdown of the Spanish edition of Google News. Note our second specification also controls for differences in long-term trends across countries. We show our results in the next section.

4. Empirical Analysis of the Shutdown of Google News Spain

4.1. Main Results

Before showing the results of our investigation for the Spanish case, we want to confirm that the shutdown of Google News in Spain did not affect the activity of Google News in other European countries. Figure 1 plots the log of daily visits for Google News webpages from Spain, France, Italy, Germany and Portugal. Note that although the jump downward in visits to Google News in Spain is clear after December 16, 2014, the number of visits to Google News does not change in the other countries.²⁶

Importantly, the shutdown of Google News in Spain appears to be an isolated event, and it did not affect other news aggregators. Figure 2 compares the log of daily visits of Google News to Yahoo! News in Spain as well as meneame.net and kiosko.net (two alternative local news aggregators). Figure 2 shows that the number of visits to these domains did not change around the time of the shutdown of Google News. Hence, the event we study here is not a confounder of major changes in Google News everywhere (Figure 1) or of changes to news aggregators in Spain in particular (Figure 2).

We now describe the overall effect of the Google News shutdown. Table 5A uses specification (1) to analyze the effects of the shutdown on the number of daily visits using in separate exercises French and Italian outlets as control groups. Columns 1 to 4 use French outlets, and columns 5 to 8 use Italian outlets as control group. The overall findings do not qualitatively change across exercises using different control groups. Columns 1 and 5 show that the shutdown decreased daily visits between 8.4% and 14.6%. See Figures 3A and 3B for a graphical representation of the negative effect of the shutdown on the number

²⁶ Despite the shutdown of Google News in Spain, the domain kept having a small number of visitors as its webpage explains Google's shutdown decision. See Figure A1 in Appendix B for the default webpage of news.google.es still posted.

of daily visits per outlet.²⁷ This finding is consistent with the fact that the market expansion effect of news aggregators outweighs the loss in daily visits due to the substitution effect.

In columns 2 and 6, we allow the effect to vary by whether the Spanish news outlet is a national, regional, business, or sports newspaper. The results show that the heterogeneity of the effect is large. Regional newspapers show a larger effect than national newspapers, business outlets show no impact, and sports outlets are always negatively affected.²⁸ These findings reflect differences in the outlets' brand awareness and brand loyalty, and are consistent with our predictions in the mechanisms section. Before the shutdown, some outlets benefitted from the visits of casual readers generated by Google News (regional and sports outlets), while others benefitted less because they experienced a smaller market expansion effect, which possibly was compensated by the switch of some (non-loyal) consumers to the aggregator (national and business outlets). These results confirm our prediction that niche outlets and predominant national outlets are those less affected by aggregators, since their specialization or ideological bias make them irreplaceable by multi-home audiences.

Columns 3, 4, 7 and 8 allow the effect to vary by ranking (top and bottom 50% within our sample according to number of daily visits), and by the share of international visitors, respectively. Consistently with the previous results, we find that the impact of the shutdown was larger in lower-ranked domains and domains with lower proportion of international visitors.²⁹ Figures 4A and 4B shows these results graphically with larger negative effects for bottom-ranked outlets.³⁰ These two findings also suggest that news outlets with more brand awareness (top 50% ranked outlets), and with a larger share of international visitors are less likely to be affected by the presence of news aggregators.

A potential concern with our analysis in Table 5A is that the Charlie Hebdo attack of January 7th 2015 and the events occurred the following days (police shot on January 8th

²⁷ Figures 3A and 3B are the result of a multi-stage process. First, we run an OLS regression of log of daily visits group-specific time trends, date and site fixed effects. Second, we calculate the error term associated with each site-date observation unexplained by time trends and fixed effects. Third, we compute the average error term per country and date. Finally, we fit a line for Spain before and after the shutdown and another line for the control group (France outlets for Figure 3A and Italian outlets for Figure 3B).

²⁸ When testing statistical differences across coefficients in column 2 of Table 5A, we are able to reject that all four coefficients are alike at the 1% level. We are also able to reject equality of any three-coefficient combination at least at the 5% confidence level. All pairwise comparison reject equality with the exception of national and business coefficients, and the regional and sports coefficients. We find the same patterns for statistical differences across parameters in column 6 when using Italian outlets as control groups.

²⁹ When testing statistical differences across coefficients in columns 3 and 4 of Table 5A, we are able to reject equality of coefficients of top and bottom-half ranked outlets at 2%, and we are also able to reject equality of coefficients of top and bottom-half international outlets at 7%. We find the same results when comparing coefficients across groups for columns 7 and 8 in Table 5A.

³⁰ Similarly to Figures 3A and 3B, Figures 4A and 4B are the result of a multi-stage process. First, we run an OLS regression of log of daily visits group-specific time trends, date and site fixed effects. Second, we calculate the error term associated with each site-date observation unexplained by time trends and fixed effects. Third, we compute the average error term per date for three distinct groups: top ranked Spanish sites, bottom ranked Spanish sites, and control group. Finally, we fit a line for Spain before the shutdown, top and bottom ranked Spanish sites after the shutdown, and another line for the control group (France outlets for Figure 4A and Italian outlets for Figure 4B).

and hostage situation at a Kosher supermarket near the Porte de Vincennes on January 9th) may be driving our results.³¹ For this reason, we repeat our analysis in Table 5B dropping all observations from January 7th to February 4th (four weeks).

Our findings in Table 5B show qualitatively consistent results with those reported in Table 5A using French news outlets as control group (columns 1 to 4 in left panel of the table). We find an overall drop in daily visits of 7.4%. This is mostly explained by a reduction of 10.9% in smaller news outlets (bottom 50% ranking), and a decrease of 9.8% in outlets with lower shares of international visitors. When we repeat this exercise using Italian news outlets as control group, we obtain qualitatively similar results. Columns 5 to 8 in Table 5B show an overall decrease of 7.9% in daily visits, and decreases of 11.5% and 10.4% for low-ranked outlets and lower shares of international visitors, respectively.³² The remarkable coincidence of results in this table across different control groups is indicative of the robustness of the result.³³

Table 6 investigates further the heterogeneity of the effect of the shutdown across outlet types. We run separate analysis for each classification using their respective French and Italian news outlets as control groups in the top and bottom panel. Specifically, we compare Spanish national outlets to French national outlets, Spanish regional outlets to French regional outlets, and so on. Columns 1 and 2 show different results for national and regional outlets depending on the control group used. On the one hand, visits to national and regional outlets decrease by 21% and 18%, respectively, when using French news outlets as control. On the other hand, visits to national outlets decrease by 8% and do not decrease for regional outlets when using Italian newspapers as control group. Columns 3 and 4 find that there is no effect on business outlets and only on sports when using Italian sites as control group. According to this set of results, Google News increased the most the number of visits from casual visitors to national, regional and sports newspapers. At this point, it is important to note that differences in the results when using French and Italian outlets as control groups may be due to differences between the news market in each country, and the different impact of the Charlie Hebdo attack in the control group.^{34 35}

³¹ Despite the fact that these events increased news consumption everywhere, they certainly increased news consumption the most in France. Hence, our results of applying diff-in-diff methodology may reflect the differential effect of the Charlie Hebdo attacks in ways that day fixed effects and a France-specific fourth-week dummy may not be able to control for.

³² In Table 5B we cannot reject equality of national and business coefficients, and regional and sports coefficients in both columns 2 and 6, and top and bottom half international coefficients in columns 4 and 8 (p-value at 12%). We can reject null hypothesis of equality for all other combinations of coefficients in the table.

³³ Results are qualitatively the same when dropping one week, two weeks or three weeks of observations after January 7th.

³⁴ We find no statistically significant difference when testing the national and regional coefficients across regressions in the upper panel of Table 6. We find statistically significant differences across all other coefficient pairs at 1%, except for business and sports at 11%. Our test for the differences in coefficients across regressions uses robust standard errors not clustered at the site level. This means that our test is conservative when not rejecting statistical difference, but not conservative enough when finding statistically significant differences if we expect standard errors to increase if clustering at the site level.

When we analyze the impact on top- and bottom-ranked outlets, we observe in columns 5 and 6 that the impact is very similar across types (14% reduction) or larger in bottom-ranked outlets (Panel B).³⁶ Finally, columns 7 and 8 consider the Top and Bottom International 50% outlets in our sample. The effect of the shutdown is larger for outlets with lower shares of internationalization (15.7% reduction vs. 11.3%) using French outlets as control group.³⁷ These findings are intuitive because internationalized outlets are less dependent on the Spanish edition of Google News to increase their brand awareness. Therefore, after the shutdown they lost less casual readers in the domestic and international market. Italian newspapers have higher shares of domestic visitors than Spanish and French outlets, and therefore we had to define a lower share of international visitor to classify them as top international. Therefore, we read the results in Panel B of Table 6 with caution because our definition may explain why we find a reduction of 10% and 7.6% for top and bottom international visitors, respectively.

We are also interested in exploring how the impact of the shutdown of Google News in Spain evolved over time until reaching steady state. For this purpose, we run specification (2) in Tables 7A and 7B for the whole sample, and for each classification separately, using French and Italian sites as control groups. Note that all specifications in Table 7A include an interaction term between the dummies "Fourth Week after 12/16/2017?" and "France" to control for the Charlie Hebdo terrorist attacks, which dramatically increased the number of daily visits to French news outlets. A common takeaway across findings in Tables 7A and 7B is that daily visits did not sharply decrease immediately after the shutdown. In each case, it took several weeks (in most cases three or more weeks) for a statistically significant decrease in daily visits to take place. While this finding may be surprising at first, this may reflect that after the shutdown casual readers remembered for some time some of the outlets they learned about with Google News. Unfortunately, we cannot test directly whether this mechanism is behind our results. However, in a related paper George and Hogendorn (2013) observe that casual readers directly visit news outlets they have discovered in the past with the use of aggregators.

Finally, we test for the "information screening" effect in Tables 8A and 8B by analyzing whether the shutdown of Google News had an impact on the consumers' engagement metrics. To do so, we use seemingly unrelated regressions that take into account correlations between our three measures of consumer engagement. Interestingly, we find a consistent long-term increase in the bounce rate across the French and Italian exercises. This increase in the bounce rate seems to be driven by a decrease in visit duration using French outlets as control group, and by a decrease in the number of pages per visit when

³⁵ When using Italian news outlets as control groups (bottom panel of Table 6), we are only able to reject equality between coefficients in columns 2 and 3 at the 9% level. All other coefficient pairs show no statistically significant difference.

³⁶ We find no significant statistical difference between coefficients across regressions in column 5 and 6 when using French or Italian outlets as control group.

³⁷ We can reject equality at the 1% confidence level, as we find a statistically significant difference between top and bottom international outlets coefficients across regressions in column 7 and 8 when using both French and Italian outlets as control group.

using Italian outlets as control group (columns 1 to 3). We find no differences in changes in engagement metrics between top and bottom ranked sites (columns 4 to 6). When we decompose the effect by week in columns 7 to 9, we find the pages per visit and the duration of visits decreased initially, but this effect vanishes for pages per visit and reverses over time for visit duration. Bounce rate consistently increases over time, with the exception of the week when Charlie Hebdo attacks occur when using France as control group in Table 8A.

To interpret this set of results in Tables 8A and 8B, bear in mind that the shutdown may have changed the composition of the consumers who visit news outlets. First, news outlets lost the search visitors who previously arrived to their web site via Google News. Second, after the shutdown, some casual consumers could substitute their visits to news aggregator for other news outlets. The findings in these tables suggest that Google News users were spending longer periods of time reading news or visiting more pages, potentially because of a better match to their interests thanks to the services of Google News. In a sense, Google News would work as screening device of news articles for its visitors who would screen articles in the front page of news outlets after the shutdown.

In summary, the results of the Spanish case reflect that the shutdown of Google News significantly reduced the number of daily visits to news outlets. The reduction of daily visits concentrated around outlets with a larger share of casual readers, such as regional and sports outlets, outlets with a low national rank and those with a relatively low internationalization level. Niche outlets and predominant national outlets were those less affected by the shutdown. This evidence leads us to conclude that brand awareness and brand loyalty are determining factors in explaining whether news aggregators play a positive role in the news market by attracting additional visitors to news outlets with low brand power.

4.2. Robustness Checks

This section presents two robustness checks that test the validity of our conclusions. First, we show an integrated analysis of the effect of the shutdown on daily visits and all consumers' engagement metrics, considering that they may be jointly determined. Second, we perform a synthetic control group analysis to assess the robustness of our results to the selection of the outlets in the control group.

Integration of results. In the previous section we have treated daily visits and engagement metrics as independent outcome variables, but it is quite plausible that these variables are jointly determined. Indeed, news outlets' visitors are not homogenous, and those that were using Google News before the shutdown could have a differentiated reading behavior that simultaneously affected their all engagement metrics. Taking this into account, we conduct seemingly unrelated regressions (sureg) with both daily visits and the consumers' engagement metrics.

Table 9 shows results of our integrated approach using Italian newspapers as control group.³⁸ Even when allowing for the correlation in the error terms of daily visits, average daily pages viewed per visit, average daily visit duration and average daily bounce rate, we still find in column 1 an 8.9% decrease in daily visits associated with the shutdown of Google News in Spain. Columns 2 to 4 show a statistically significant decrease in the average daily number of pages viewed per visit (4.9% decrease) and a statistically significant increase in the average daily bounce rate (0.4% increase). We find no impact on the average daily visit duration.

When investigating differences between top half and bottom half ranked sites, results in columns 5 to 8 show a larger impact on daily visits of bottom-half ranked outlets than on top-half ranked outlets (11.8% versus 3.3%). Yet, we find no significant differences on impact for average daily pages per visit (5.3% versus 4.1%), visit duration (non-significant 2.7% versus 2.6%), and bounce rate (0.44% versus 0.46%).

Finally, columns 9 to 12 investigate the impact over time of the shutdown of Google News on our four outcome variables while allowing for correlation in the error term. Results after the 7th week of the shutdown are consistent with our findings in columns 1 to 4. If anything, we observe gradual increases in visit duration 4 to 6 weeks after the shutdown. In the end, the long-term impact on visit duration is null.

Selection of the control group. A second related concern is our choice of French and Italian news outlets as separate control groups for Spanish news outlets. To examine the robustness of our results to the selection of the outlets in the control group, we perform a synthetic control group analysis with data from Germany, France, Italy and Portugal (see Table A1 in Appendix B for a list of Portuguese news outlets).³⁹ Specifically, with this method we consider a weighted average of control news outlets (synthetic control) that is as similar as possible to the treated Spanish news outlets regarding the pre-treatment outcome variable. The benefit of building this synthetic control group is that the pre-shutdown characteristics of the Spanish news outlets can be better approximated by a combination of untreated news outlets than by an unweighted group of outlets (Abadie and Gardeazabal, 2003; Abadie et al. 2015).

In order to implement this analysis, we collapse our outlet-day specific data into group-week observations where we define our groups by country of origin (Spain, France, Germany, Italy or Portugal). This group classification allows us to create a synthetic control group for the average Spanish news outlet and 4 potential control groups. The synthetic control group method optimally weighs the outlets of the control group to match the behavior of the outlets of the treatment group before the shutdown of Google News. While

³⁸ Table A3 in Appendix B shows qualitatively similar results when using French news outlets as control group.

³⁹ Because Portuguese news outlets are much smaller in number of daily visits than Spanish, French and German outlets, the data from SimilarWeb on Portuguese outlets is noisier and less precise than for outlets from other countries.

the weights are fixed after the shutdown, the visits to the outlets in the control groups change over time and so will the synthetic control. This is advantageous because it accounts for the effects of the confounders changing over time (unlike regular difference-in-differences methods). To match behavior between treated and control groups prior to treatment, this method creates optimal weights using the number of daily visits (for only a subset of the pre-treatment period), share of domestic visits, national rank, and engagement metrics.

Figures 5A and 5B show the average daily visits for Spanish news outlets and its synthetic counterpart during the period analyzed. While the exercise in Figure 5A considers outlets from all four countries as potential controls (France, Germany, Italy and Portugal), the analysis in Figure 5B considers only Italian and Portuguese outlets to avoid the distortions generated by the Charlie Hebdo attack in France and the Google disruption in Germany in October 2014. Note that in both figures, the synthetic control closely tracks the average Spanish news outlets prior to the Google News shutdown. After the shutdown, the average Spanish outlet starts to diverge from the synthetic control unit and is consistently below the synthetic control group.

A second exercise only considers national news outlets across Germany, France, Spain, Italy and Portugal. Similarly to our previous exercise, we collapse our outlet-day specific data into group-week observations where we define our groups by country of origin. Figure 5C shows the results of using all four countries as potential control groups, and Figure 5D shows results using national news outlets from only Italy and Portugal. Note that in both cases the treatment and control closely track each other prior to the shutdown of Google News, and they diverge after the shutdown. These results show that our findings in Tables 5A and 5B are robust to the use of synthetic control group methodology.

4.3. The Market for Advertisement

This section complements our assessment of the shutdown of the Spanish edition of Google News with an analysis of its effects on advertisement revenues. While our findings in previous sections show that the shutdown reduced the volume of daily visits received by news outlets, we next examine whether it financially affected the outlets by decreasing the advertisers' spending and their advertising strategies. We expect the reduction of visits in news outlets to reduce their ad inventory and therefore their advertisement revenues. Moreover, the shutdown could also affect advertisers' expected internet sales and their demand for advertising slots.

We use daily data on advertising metrics at the domain-advertiser-page level from Arce Media, a firm specialized in the collection and analysis of advertisement information in Spain. ARCE Media collects daily advertising information for a sample of websites that commercialize a large ad inventory. For the purpose of our analysis, we separate these websites into two groups, namely, news outlets and non-news outlets. Most of the non-news outlets obtain their main source of revenues from activities other than advertising

(e.g. ebay), but all of them are big players in the advertising market and compete with news outlets for advertisers. From an empirical point of view, we want to clarify two points. On the one hand, we use site fixed effects in our regression specifications, and therefore differences in advertising levels among websites are controlled for. On the other hand, Google News certainly did not index non-news sites and so this control group is an optimal candidate when studying the effect of the shutdown on advertising revenues. While both news and non-news sites compete for advertisers, only news sites were affected by the shutdown of Google News.

The advertisement data contains information for 78 online domains, 47 news outlets and 31 non-news outlets (Table A2 in Appendix B provides a full list of the outlets in our sample). We consider several measures for our analysis. *Advertisement Intensity* is a variable that reflects the intensity of the advertisers' campaigns. Arce Media visits several times per day the web site of the news outlet (and of the other domains), and for each of them it calculates the number of times an advertiser appears in the news outlets. *Daily Revenue* measures the estimated daily advertisements revenues obtained by a news outlet in its front page and in its content pages. Arce Media calculates this variable taking into account the advertising intensity of each advertiser in the outlet, the number of daily visits, and the prices charged by the news outlet for each type of advertisements. Once we know the Advertisement Intensity and the Daily Revenue per outlet, advertiser, day and front page level, we can calculate the ratio between these measures, and we can also aggregate the information at the day level per advertiser-site pair. We can also collapse the data at the domain and front page level to account for the number of Daily Advertisers that promote their products in the news outlets.

The price paid by advertisers to news outlets usually depends on the cost per thousand impressions or per page views (CPM), which is the expense incurred for every thousand potential customers who view the advertisement. These prices are either negotiated directly with advertisers by sites or sold to an intermediary “Ad Network” who distributes the remaining ad inventory in the market at bulk (see Appendix A for a thorough description of the Spanish advertising market). Taking this into account, we expect the shutdown of Google News to reduce the outlets' advertisement revenues, because of either a decrease in CPM (prices will be lower if the demand for slots and the marginal revenue curve decreases), a decrease in advertising intensity or a general decrease in visits or impressions. Similarly, we would expect this decrease to be weaker in front pages than in other pages because the links of news aggregators direct consumers to pages other than the front page.

We follow the diff-in-diff methodology used along the paper with two main differences. First, we use Spanish news outlets as the treatment group and Spanish non-news outlets as the control group. Second, we only use data from December 2014 because Arce Media has reported a change in their methodology to collect their data after January 2015. For that reason, and given the granularity of the data at the site-advertiser-day-page level, we focus on changes occurred in advertising behavior at the site-advertiser-page level within two weeks before and after the shutdown of Google News.

Table 10 shows our first set of results using all Spanish news outlets as treatment group, and non-news outlets as control group. Columns 1 to 3 show results of using all data (front page and other pages) with site*advertiser*front page and date fixed effects with group specific trends. Columns 4 to 6 only use front page data and site*advertiser fixed effects, while columns 7 to 9 use only data from other pages (not front page) together with site*advertiser fixed effects. The results in this table show consistently that daily revenues and daily revenues per advertising intensity unit decrease after the shutdown of Google News. As predicted, the advertising intensity did not change in front pages and decreased in other pages.

We next aggregate data at the site-day-page level, and run diff-in-diff specifications in Table 11. Specifications reported in Panel A (top panel) use all observations (front page and other pages) with site*page and day fixed effects. We find a decrease in the number of daily advertisers, the revenue per advertiser and page, and the revenue per advertising intensity after the shutdown. Panels B and C use observations from front pages and other pages, respectively, with site and date fixed effects. We find decreases in revenue, ad intensity, number of daily advertisers, average revenue per advertiser and revenue per advertising intensity in front pages, and decreases in ad intensity and ad intensity per advertiser in other pages.

Differences in results between Table 10 and 11 reveal differences in the effect of the shutdown at the advertiser*site level and at the aggregated site level. These differences may reflect a heterogeneous impact of the shutdown across news outlets and advertisers.⁴⁰ On the one hand, Table 10 shows that intensity for those that advertise before and after the shutdown (advertiser-site fixed effects are in place) went down overall, but it went down more so in non-front pages (4%) and not at all in front pages. These results make sense taking into account that (i) news aggregators link the pages of the news stories and not the front page of the outlets, and (ii) willingness to pay for ads in front pages is larger than in non-front pages. On the other hand, Panel A of Table 11 shows that intensity did not go down overall, but at the same time we do observe statistically significant decreases in the number of daily advertisers (which explains differences with Table 10), revenue per advertiser, and revenue per intensity unit. Panels B and C also show that the intensity per advertiser did not go down in front pages, but it did in non-front pages (consistent with Table 10). Daily intensity went down in both front and other pages, number of advertisers went down in front pages but not in non-pages, and revenue mostly went down in front pages and it did not in non-front pages. In our opinion, these results are overall consistent when comparing advertising intensity at the advertiser level or intensity divided by number of advertisers at the site level. While we do not observe advertising prices, a possible decrease in prices for non-front page advertising could help to interpret these findings.

⁴⁰ Tables A4 and A5 in Appendix B shows results of running seemingly unrelated regressions (sureg). While the former runs sureg with advertising intensity and revenue per advertising intensity as dependent variables at the site-advertiser-page type-day level, the latter does so for advertising intensity, number of daily advertisers, advertising intensity per advertiser, and revenue per advertising intensity at the site-day level and at the site-page type-day level. These specifications allow for the error term to be correlated across regressions.

In a nutshell, our results indicate that the shutdown of Google News was not innocuous for news outlets from the advertising market perspective. The reduction in daily visits decreased the advertisers' spending for advertising slots, reducing advertising intensity mostly on other pages while remaining constant in front pages. At the aggregate site level, this implied a reduction in revenues, advertising intensity and the number of advertisers. We also find statistically significant reductions in the revenue per advertising intensity as well as the revenue per advertiser in front pages, and reductions in the advertising intensity per advertiser in other pages.

Finally, it is important to note that a potential limitation in the interpretation of our results is that we are unable to observe the underlying heterogeneity in the use of direct and programmatic advertising by news outlets. In Appendix A, we detail that even though programmatic advertising is sizable nowadays, at the time of the Google News shutdown, only a relatively small part of the ad inventory was commercialized through programmatic advertising. Yet, the presence of targeted advertising poses the question of how advertisers using this option reacted to the shutdown of Google News, and whether programmatic advertising allows for a quicker adjustment of the advertising campaigns to the varying conditions of the market. Our analysis in this section cannot separate which part of the effect of the shutdown was due to adjustments in direct and programmatic advertising.

5. The Opt-in Policy in Germany

This section studies the impact of Google's opt-in policy in Germany. Two major differences exist between the German case and the previously analyzed Spanish case. First, the mechanisms at work are different. In Germany, even though Google News continued to index all news outlets after the adoption of its opt-in policy, it gave a different treatment to the outlets that opted-out. Specifically, the aggregator complemented the links to the outlets opting out with shorter excerpts and it could not use images. We call the "competitive effect" the impact of the difference in information provided by Google News for links of outlets that opted out relative to the information provided for links of outlets that opted in. Notice that in Germany the outlets that opted out still experienced the substitution effect from Google News, as the aggregator did not shutdown. However, it is possible that they did not completely benefit from the market expansion effect if the traffic that they could potentially receive from the aggregator ended up in the outlets that opted in (Dellarocas et al. 2016; Huang, 2017; Jeon, 2018). In this section, we use this case to examine the role that the information portrayed in the links' excerpts plays in the consumers' decision to click through the links, and we discuss the consequences of this "competitive effect".⁴¹

⁴¹ Note that the German case does not allow us to identify separately the "competitive effect" from the "information screening effect" introduced in the mechanisms section. Briefly, the adoption of the opt-in policy did not prevent Google from indexing news from opted out domains (although it had to reduce the length of the excerpts) and therefore some Google users were able to substitute the outlets that opted-out for others that opted-in.

The second relevant difference between the Spanish and the German case is that in the latter case the treatment period we examine took place for a finite amount of time from October 23 to November 5, 2014.⁴² After this period, VG Media outlets decided to opt back in to Google News. The short opting-out period means that we are not able to estimate the long-term consequences of the different treatment that Google News gave to the outlets that opted-out.

5.1. Empirical Methodology

We analyze the impact of VG Media's decision to opt out from the Google's policy, by comparing German, Italian and French news outlets during the treatment period. Our first specification compares German news outlets (treated group) with French and Italian news outlets (control groups) before, during, and after the de facto opt-out period from October 24 to November 5, 2014 (treatment period). It is as follows:

$$\ln[y_{ijt}] = \theta \text{Germany}_i * \text{Opt_Out}_t + \mu_{j0}t_{jt} + \mu_{j1}t_{jt}^2 + \gamma_i + \delta_t + u_{ijt},$$

where all dependent variables are defined as in the previous section. Germany_i is a dummy that equals 1 if online newspaper i is German, and 0 otherwise. The dummy Opt_Out_t takes the value of 1 if day t is between October 24 and November 5, 2014, and 0 otherwise. On the other hand, t_{jt} and t_{jt}^2 are group-specific time trends. Finally, γ_i and δ_t are country-site specific fixed effects and date fixed effects that control for unobserved time-invariant country-site-specific factors and date-specific factors common to all sites, respectively. We control for differences in time trends to take into account long-term differences across groups that are not captured by our date fixed effects, and that could be mistaken by effects of the Google opt out period under study. We assume the error term u_{ijt} to be iid and normally distributed as usual.

We also consider other specifications in which the treated group is the outlets that opted out. Specifically, we run separate regressions for the members of the VG Media association and for the group of news outlets controlled by Axel Springer. Axel Springer was one of the most active publishers in advocating for a change in German copyright law and was the first to announce its opt-out choice in October 2014. Finally, we also break the Opt_Out_t dummy into $1stWeekOpt_Out_t$ and $2ndWeekOpt_Out_t$ that take the value of 1 if day t falls in either the first or second week, respectively, of the full opt-out period. A finding of a negative θ in these specifications would imply that Google News generated a larger market expansion effect in the outlets that opted-in than in those the opted-out, due to the “competition effect.”

⁴² Notice that the German event took place before the Google News shutdown in Spain. We present the Spanish case first for expositional reasons.

5.2. Results

We begin the analysis of the German case by estimating the effects of the opt-out decision on the outlets' daily visits using Italian outlets as control group. Columns 1 in Table 12 shows that after the opt-out decision, the number of daily visits to German outlets did not change relative to Italian outlets. Column 2 shows that this finding is robust when splitting German outlets into VG Media sites and non-VG Media sites. If anything, column 2 shows that the number of daily visits to VG Media outlets experienced a statistically non-significant decrease of 2.4%. We focus on Axel Springer sites in columns 3 and 4, and our results show that while the number of daily visits did not seem to change in German outlets overall, the number of daily visits went down in Axel Springer sites by 7.6%. This finding is robust when we separate the effects by weeks: the number of daily visits went down in both weeks of the opt-out period (9.7% and 7.4% in the first and second weeks, respectively).

The rest of our specifications in Table 12 use German outlets only, and considers the effect in the daily visits on the news outlets that opt out using as a control group the rest of German outlets in our sample. Column 5 shows again the existence of a negative but non-significant effect of the opt-out decision on the visits to the VG Media outlets relative to all other German outlets in our sample. In columns 6 and 7, we focus on the 10 outlets of our data set that Axel Springer controlled during this time. We again find a negative and significant reduction in daily visits of around 7.6% in Axel Springer outlets relative to all other German outlets in our data. This effect was stable across weeks during the treatment period, with an 8.1% and 8.9% decrease in the first and second weeks of the opt-out period, respectively. To summarize, our results suggest the change in Google's linking policy created a “competitive effect” that diverted some of the page views of the outlets that opted out to those that opted in. Columns 1 and 2 show that the opting out of VG Media outlets did not significantly reduce the number of news stories read by German consumers, and columns 3 to 7 reveal that the “competitive effect” was heterogeneous across the outlets that opted-out, being only significant for the outlets related to Axel Springer. Notice that these outlets are on average higher ranked, and received more daily visits than other VG Media's outlets.⁴³

A reasonable concern when interpreting our findings of columns 6 and 7 in Table 12 is the endogeneity surrounding the decision of opting out by VG Media and Axel Springer outlets. To address this issue, column 8 shows the results of running propensity score matching between all Axel Springer outlets and the rest of German outlets in our sample as control group, and running difference in differences regressions using the closest match as control group for each treated Axel Springer outlet. Column 8 shows a decrease of 8.1% in

⁴³ Axel Springer is one of the largest publishing houses in Europe and one of the main contributors of outlets to our sample of German sites, the largest within the conglomerate VG Media with ten outlets out of sixteen. The six VG Media's outlets not part of Axel Springer are regional outlets with a low domestic ranking, business and sports outlets.

daily visits consistently with other findings in columns 6 and 7.⁴⁴ We find qualitatively similar results in Table A6 in Appendix B when using French news outlets as control group.⁴⁵

Next, Table 13 performs a robustness check to determine whether the Axel Springer dummy variable may be capturing the impact of shocks on the demand for news that only affected specific outlet types during the treatment period. To do so, we repeat the analysis for the Axel Springer outlets in columns 6 and 7 of Table 12 while also including as independent variables the interaction between the *Opt_Out* dummy variable and the specialization categories of outlets. Columns 1 and 2 take into account the classification of the news outlets according to their content (national, regional, business, and sports), and columns 3 and 4 according to their ranking (top 50%). Results in columns 1 and 2 show the effect of the opt-out decision was only significant in the second week of the treatment. However, the estimates in columns 3 and 4 offer similar insights to those in Table 12.⁴⁶

In summary, our analysis of Google’s opt-in policy in Germany provides a set of interesting results. We have shown that the VG Media’s decision to reject the agreement with Google News had an overall negative but non-significant effect of their daily visits. Moreover, the effect of this decision was heterogeneous across VG Media’ outlets, since only the 10 outlets in our sample controlled by Axel Springer experienced a significant average reduction in daily visits of around 8%. This explains why Axel Springer and other VG Media outlets finally decided to accept the conditions of Google News, and ended up opting back in. These results are overall consistent with the “competition effect” identified in other recent works. On the one hand, Jeon and Nasr (2016) and Jeon (2018) study the factors that induce news outlets to be indexed by news aggregators. They show that if the third-party content indexed by the aggregator generates more traffic to each outlet, then outlets will decide to opt in, and that their interest in being indexed will increase with the size of the aggregator’s third-party content. On the other hand, Dellarocas et al. (2016) study how readers allocate their attention to different article links within a Swiss news aggregator. They find that having longer than average excerpts and the presence of accompanying images increase the probability of links of being chosen by consumers. In this context, Dellarocas et al. (2016) explain that the “publisher’s unilateral decision to shorten the snippet lengths of its articles and/or disallow the reproduction of images might put them at disadvantage in situations where there are several related articles on the same topics”.

⁴⁴ We use kernel propensity score matching through the “diff” command in Stata/MP 15.1. We estimate the propensity score with the variables National, Regional, Business, Sports, National Rank and % Domestic Visits. We perform the diff-in-diff only on the common support of the propensity score with the kernel option.

⁴⁵ If anything, columns 1 to 4 in Table A6 of Appendix B finds that German outlets on average received higher number of daily visits than French news outlets. In our opinion, this result may be due to the sharp increase and posterior decrease in daily visits due to the Charlie Hebdo attack relative to the rather stable profile of German outlets in 2015. In any case, we also find that Axel Springer outlets experience a decrease of around 8% relative to other German outlets.

⁴⁶ Table A7 in Appendix B examines the effects of Axel Springer's opt-out decision on the engagement metrics. We do not observe statistically significant changes.

Our empirical findings are in line with the predictions of these papers. We find that those news outlets that opted out suffered from traffic loss because they could not completely benefit from the aggregator's market expansion effect. Even if news outlets may collectively prefer short excerpts to increase the likelihood that consumers click their links, short excerpts cannot be sustained as an equilibrium when news outlets can deviate and accept longer excerpts to become more attractive. The evaluation of the German case complements our earlier results from the Spanish experience. The analysis of the shutdown of Google News in Spain is important to understand the net effect of market-expansion and substitution effects of news aggregators, in a context where there were no "competition effect" as all news outlets received the same treatment. In contrast, the analysis of the German case studies an event where Google News did not shut down, but it gave different treatments to different groups of outlets. Therefore, Google News still generated a substitution effect on all news outlets, but created different market-expansion effects for different groups of outlets. The source of differences in the market-expansion effect across outlet groups is the "competition effect" identified in our analysis.

Note that the similar magnitude of the "competition effect" in Germany and the net effect of the shutdown in Spain (around 8% in both cases) suggests that the size of the substitution effect can be modest. A potential explanation for this finding, which we cannot test for with our data, may be that consumers multi-home and use aggregators to gain access to additional contents than those offered by their reference news outlet. Our results in the German case also reveal the important role of excerpts in news aggregators and their capacity to modify consumers' reading behavior. Understanding how the characteristics of excerpts can transform aggregators in a complementary channel for news outlets or in a substitute service is a puzzle for future work.

6. Limitations and Concluding Remarks

Amidst the growing importance of online platforms, news aggregators are one of the most successful new players in the internet's new era, quickly rising to occupy top positions in audience rankings. Yet, since their introduction, they have faced the opposition of news publishers, who consider aggregators as free-riders that resell their content. This controversy has motivated the amendment of copyright laws in several countries, which have limited the use aggregators can make of the publishers' content. Google News' strategy in this new environment has been to avoid paying any link fee for the indexation of news stories. In Spain, after the government created a compulsory link fee, Google News shut down its Spanish edition, sending a clear message to the publishers and governments of other countries that it would not accept paying for indexing news stories. In Germany, where the linking fee could be negotiated, Google has adopted an opt-in policy that, in practice, forces news outlets to waive any linking fee. Google is complementing this strategy with other actions, such as the creation of the Digital News Initiative, which gives support to European publishers for developing products that increase their revenue and traffic, stimulate innovation in digital news journalism, and promote training and academic research into journalism. These solutions seem to have left unsatisfied most traditional publishers, who in the last decade have seen how the significant increase in their online

visits and advertising revenues have not compensated for the reduction in advertisement revenues of printed newspapers. Consequently, the debate about regulating this market continues, and in the last year, European publishers have managed to move the discussion from the national arena to the European Union level. The European Union will have to decide whether to include this issue on its revision of the copyright legislation.⁴⁷

The goal of our research has been to examine the role of news aggregators and their effect on the number of visits and the advertising revenues of online news outlets. The economics literature has identified two potential types of effects news aggregators may have on news outlets. Whereas aggregators facilitate indirect visits of casual readers to news outlets that otherwise would not take place (market expansion effect), news outlets also compete with aggregators for the direct visits of non-loyal readers, who are aware of the outlets but that may prefer the aggregators' screening capacity (substitution effect). Our main contribution to the existing literature is to estimate the relative strength of these two effects and to show how the news outlets consumers' base can determine the benefits they obtain from the aggregators.

Our analysis of the shutdown of the Spanish edition of Google News shows a significant reduction in the audience of news outlets. This result confirms that news aggregators are an important channel for attracting visitors to news outlets. Our findings that outlets with smaller brand power benefit the most from news aggregators suggest that aggregators are not only a mechanism to screen news stories, but also allow consumers to recall and discover new sources of information while improving their access to diversified contents. We also show that the shutdown of Google News in Spain reduced revenues and advertising intensity of advertisers and it did so more intensely on non-front pages.

Our analysis of the German case has shown that changes in the size of the excerpts or the images the aggregators release modify the traffic news outlets receive (competition effect). In Germany, Axel Springer's decision to opt out from Google News significantly reduced the number of daily visits received by their outlets. Moreover, the traffic these outlets could have received from Google was diverted to other outlets that opted in, which reveals the relevance of news aggregators for competition.

Limitations. Despite the robustness of our main findings, our study is not free of limitations. On the one hand, our domain-level data does not allow us to separately identify the market expansion and substitution effects of news aggregators. Similarly, our sample does not contain the universe of news sites or information on the number of visitors through mobile devices. While selecting from the top of the distribution of news outlets may yield a lower bound of the total effect (if in fact lower ranked outlets benefit the most from news aggregators), the lack of data on mobile visits may magnify the impact of news aggregators.

Additionally, our revenue data does not contain information on advertising prices. While we restrict our advertising analysis to 15 days before and after the shutdown, we cannot

⁴⁷ See https://elpais.com/elpais/2017/03/24/inenglish/1490362000_295766.html.

truly disentangle whether differences in advertising revenues per advertiser in a site are due to changes in prices or the number of daily visits.

The use of French outlets as control causes concern due to the impact of the Charlie Hebdo attacks on the number of daily visits to French news domains. We attenuate this concern using Italian outlets as control group in a separate exercise, and yet Italy's profile of internet news consumption is not as good of a match to Spanish internet profile as France's profile is.

Finally, our analysis of the German case relies on a very short period of time (namely two weeks) and the self-selection of VG Media and Axel Springer sites onto the opt-out program. This calls in for the question of whether the competition effect is then lessened because Axel Springer sites were among the most important sites to begin with. Future studies of the importance of news aggregators should aim to address these issues.

Conclusions and Policy Implications. In summary, our research finds that news aggregators benefit news outlets, and they do even more when news outlets have a small brand power. If anything, our results show that on average no outlet is negatively affected by aggregators. As a matter of fact, outlets benefit from news aggregators both through an increase in traffic and increases in other performance indicators such as advertising revenues, advertising slots, and the number of advertisers. This set of findings has very important implications for policy makers interested in understanding the impact that could have a new copyright legislation limiting the aggregators' access to the publishers' contents.

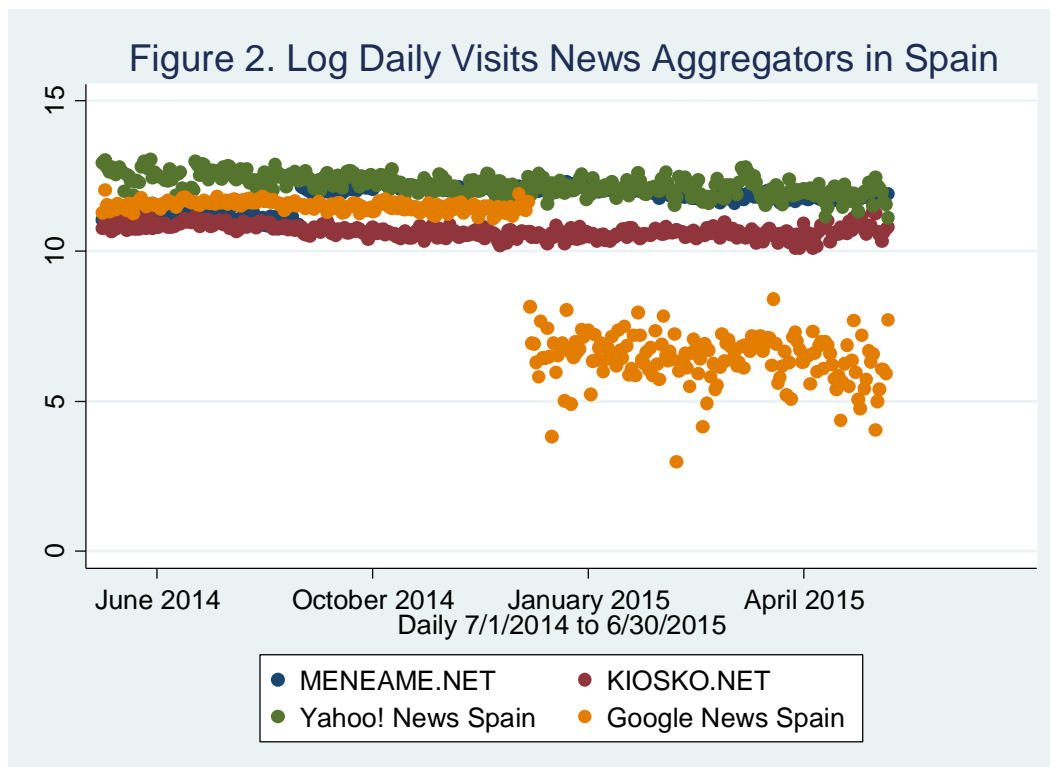
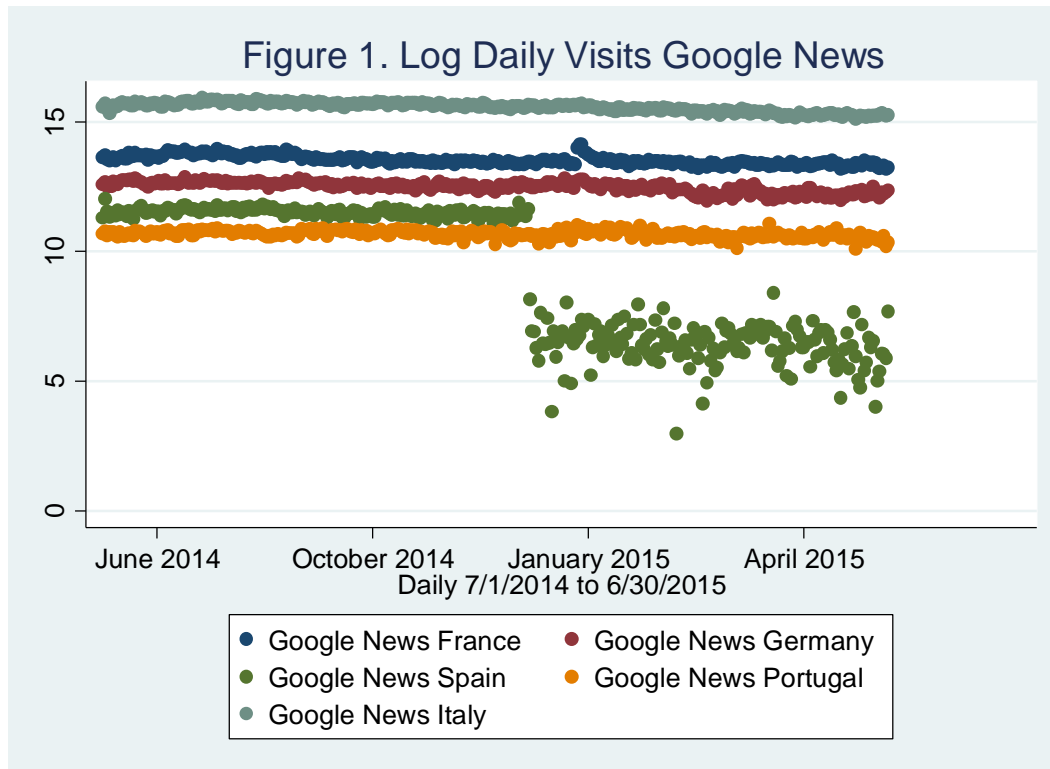
Although our research answers the question of whether news aggregators are predominantly increasing the audience and market reach of news outlets, we believe future research should further examine the impact of news aggregators on consumers' engagement metrics. A related question that merits additional investigation is whether news aggregators affect the content quality of news outlets and the composition of their readership through their impact on the advertising market.⁴⁸ A full understanding of the effect of news aggregators on news consumption is essential for copyright policy design that benefits consumers and societies overall.

⁴⁸ In a recent paper, Angelucci and Cagé (2017) analyze the impact of the introduction of advertisement on television in 1968 on French news outlets. They find that the change in the advertisement market lead to a decrease in advertising revenues of national newspapers compared to local newspapers. This shock led national newspapers to lower their advertising price, and also their subscription price relative to local newspapers. Moreover, national newspapers reacted by producing less journalistic-intensive content. Hence, they find a positive relationship between advertising revenues and quality of information. Angelucci et al. (2018) study the impact of entry of TV broadcast national news in the 50s in the US news market.

References

- Abadie, A. and J. Gardeazabal (2003). "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, Vol. 93(1), 113-132.
- Abadie, A., A. Diamond and J. Hainmueller (2015). "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, Vol. 59(2), 495-510.
- Angelucci, C. and J. Cagé (2017). "Newspapers in Times of Low Advertising Revenue," mimeo.
- Angelucci, C., J. Cagé and M. Sinkinson (2018). "Media Competition and News Diets," mimeo.
- Athey, S., E. Calvano and J. Gans (2011). "The Impact of Internet on Advertising Markets for News Media," Discussion paper, Harvard University.
- Athey, S. and Mobius, M. (2012). "The Impact of News Aggregators on Internet News Consumption: The Case of Localization," Working Paper, Microsoft Research.
- Athey, S., M. Mobius and J. Pal (2017). "The Impact of Aggregators on Internet News Consumption," Working Paper, Microsoft Research.
- Cagé, J., N. Herve, and M.-L. Viaud. (2015). "The Production of Information in an Online World," NET Institute Working Paper #15-05.
- Calzada, J. and Ordóñez, G. (2012). "Competition in the News Industry: Fighting Aggregators with Versions and Links," NET Institute Working Paper No. 12-22.
- Chesnes, M., W. Dai, and G. Z. Jin (2017). "Banning Foreign Pharmacies from Sponsored Search: The Online Consumer Response," *Marketing Science*, 36(6): 879-907.
- Chiou, L., and C. Tucker (2017). "Content Aggregation by Platforms: The Case of the News Media," NET Institute Working Paper No. 11-18. Forthcoming in *Journal of Economics and Management Strategy*.
- CNMC (2014). "Propuesta referente a la modificacion del articulo 32.2 del proyecto de ley que modifica el texto refundido de la Ley de Propiedad Intelectual," Comisión Nacional de los Mercados y la Competencia; PRO/CNMC/0002/14.
- Dellarocas, C., Katona, Z. and Rand, W. M. (2013). "Media, Aggregators, and the Link Economy: Strategic Hyperlink Formation in Content Networks," *Management Science*, 59(10): 2360-2379.
- Dellarocas, C., J. Sutanto, M. Calin, and E. Palme. (2016). "Attention Allocation in Information-Rich Environment: the case of Aggregator," *Management Science* 62, 9, 2543-2562.
- George, L. and Hogendorn, C. (2012). "Aggregators, search and the economics of new media institutions," *Information Economics and Policy*, 24, issue 1: 40-51.

- George, L. and Hogendorn, C. (2013). "Local News Online: Aggregators, Geo-Targeting and the Market for Local News," Working paper.
- Goldfarb, A. and C. Tucker (2017). "Digital Economics," NBER Working Paper No. 23684.
- Huang, J. (2017). "Should Google Charge Like a Taxi Driver?," CEMFI Working Paper.
- Isbell, K. (2010). "The Rise of the News Aggregator: Legal Implications and Best Practices," Berkman Center Research Publication No. 2010-10.
- Jeon, D.S. and N. Nasr (2016). "News Aggregators and Competition Among Newspapers in the Internet," *American Economic Journal: Microeconomics* 8(4): 91–114..
- Jeon, D.S. (2018). "Economics of News Aggregators," in *Economic Analysis of the Digital Revolution, FUNCAS Social and Economic Studies, 5*, edited by Juan Jose Ganuza and Gerard Llobet, FUNCAS, Madrid, Spain.
- Llobet, G. (2015). "Los agregadores de noticias y la prensa escrita en la era de Internet," *Papeles de Economía Española*, N. 145, 172-179.
- NERA (2015). "Impacto del Nuevo Artículo 32.2 de la Ley de Propiedad Intelectual Informe para la Asociación Española de Editoriales de Publicaciones Periódicas (AEEPP)," <http://www.aepp.com/pdf/InformeNera.pdf>
- NERA (2017). "Impact on Competition and on Free Market of the Google Tax of AEDE fee: Report for the Spanish Association of Publishers of Periodical Publications (AEEPP)," http://www.aepp.com/pdf/Informe_NERA_para_AEPP_%28INGLES%29.pdf
- Roos, J., C.F. Mela and R. Shachar. (2015). "The Effect of Links and Excerpts on Internet News Consumption," mimeo.
- Rutt, J. (2011). "Aggregators and the News Industry: Charging for Access to Content," Working Papers 11-19, NET Institute.
- Sismeiro, C. and A. Mahmood (2018). "Competitive Versus Complementary Effects in Online Social Networks and News Consumption: A Natural Experiment," *Management Science*.



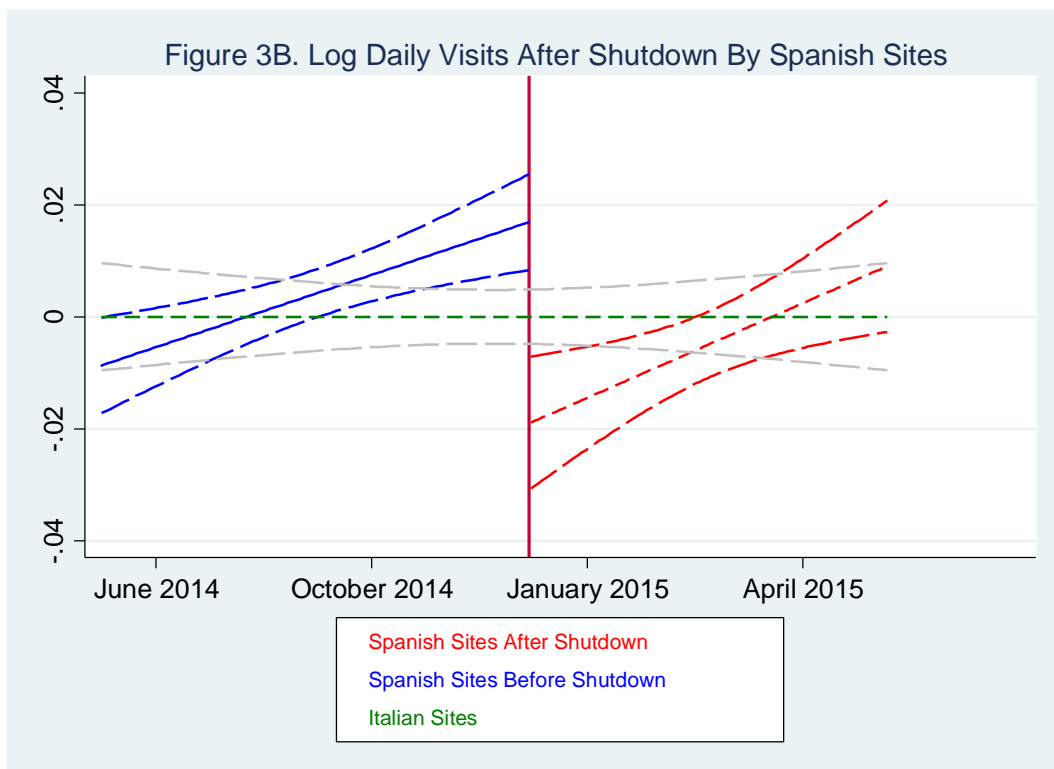
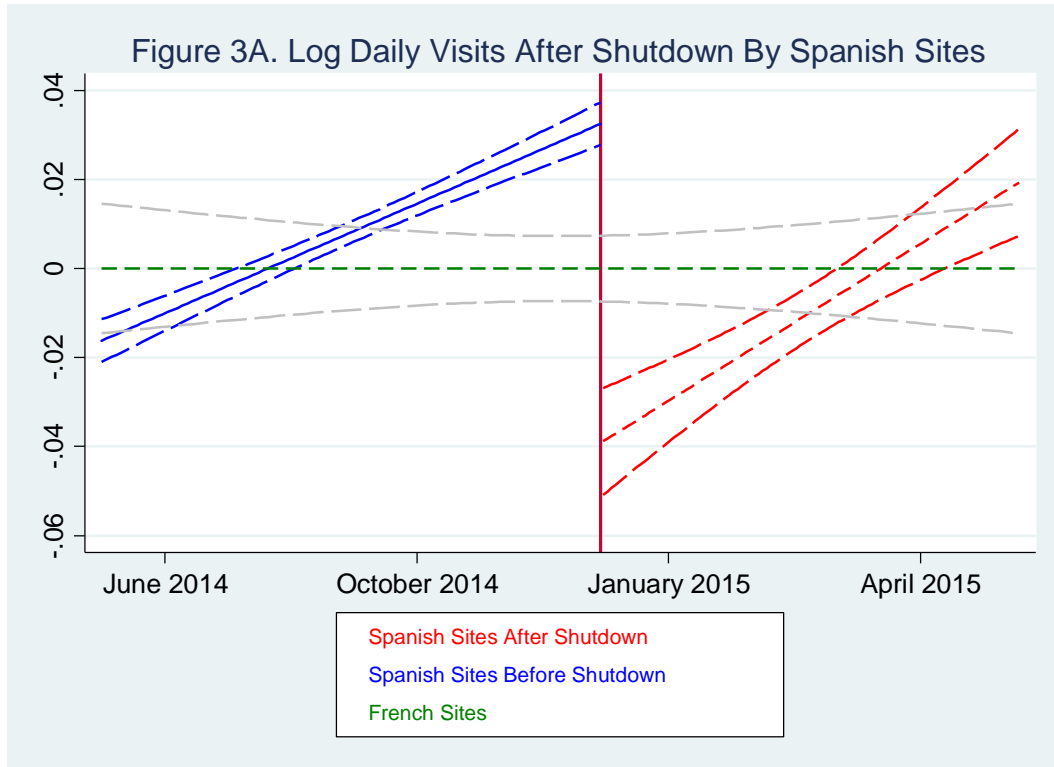


Figure 4A. Log Daily Visits By Top/Bottom Ranked Site

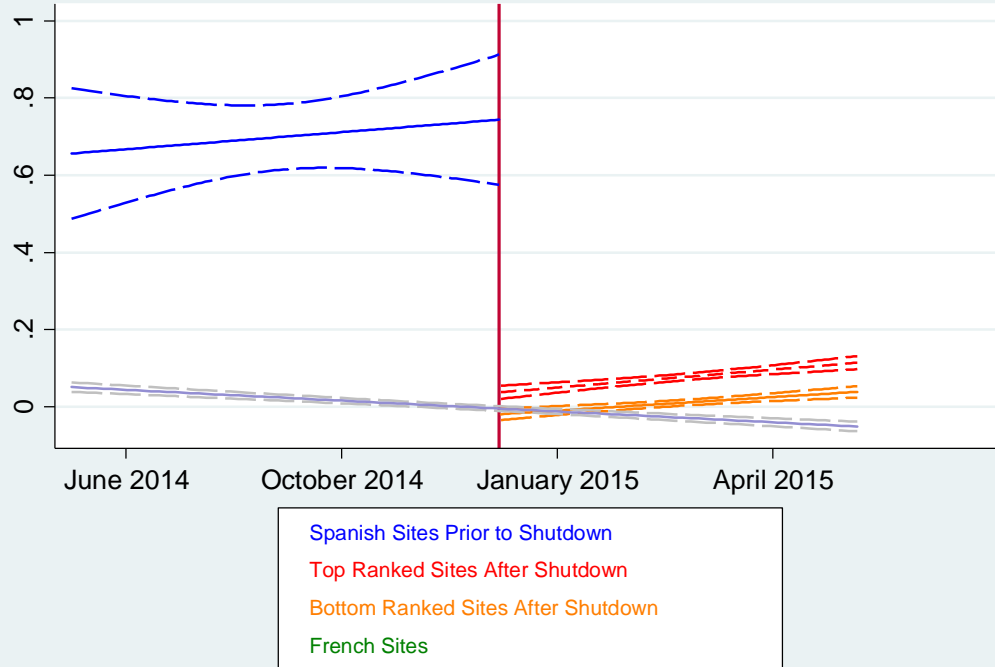


Figure 4B. Log Daily Visits By Top/Bottom Ranked Site

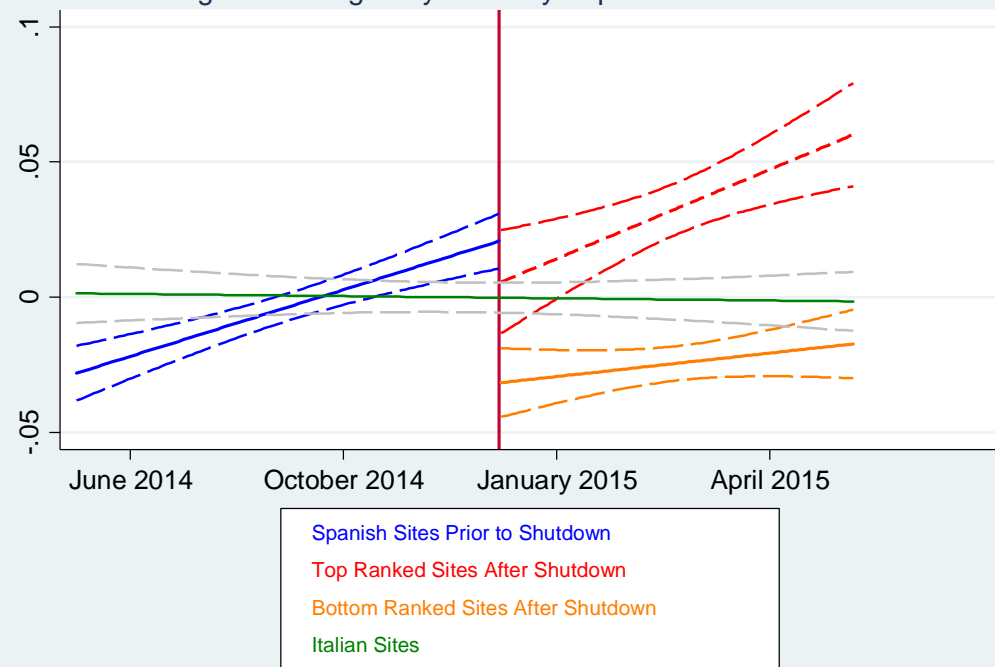


Figure 5A. Synthetic Control Analysis All News Outlets
Using France, Germany, Italy and Portugal News Outlets

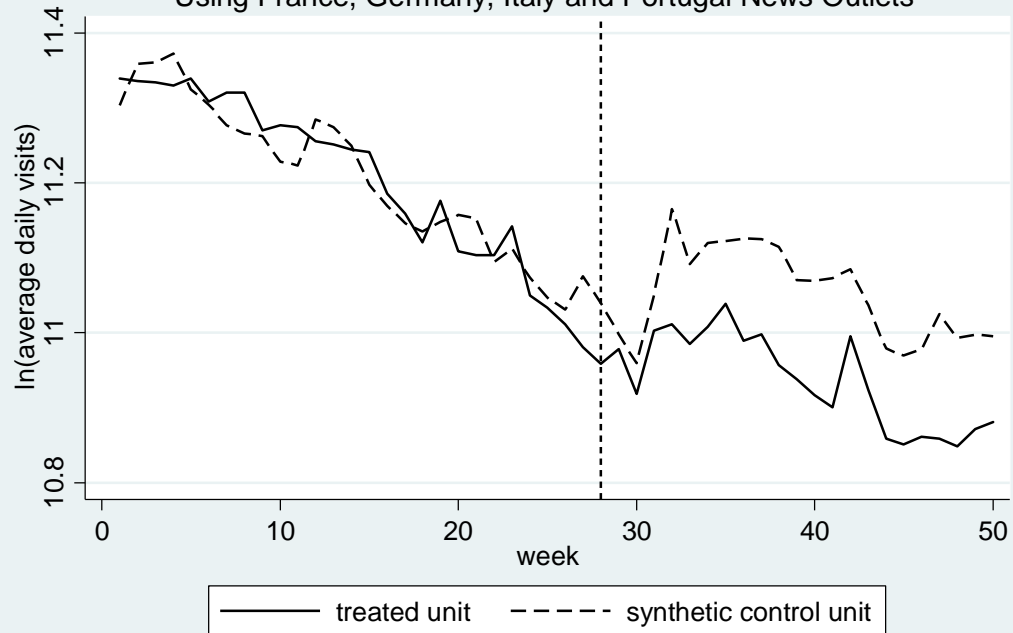


Figure 5B. Synthetic Control Analysis All News Outlets
Using Italy and Portugal News Outlets

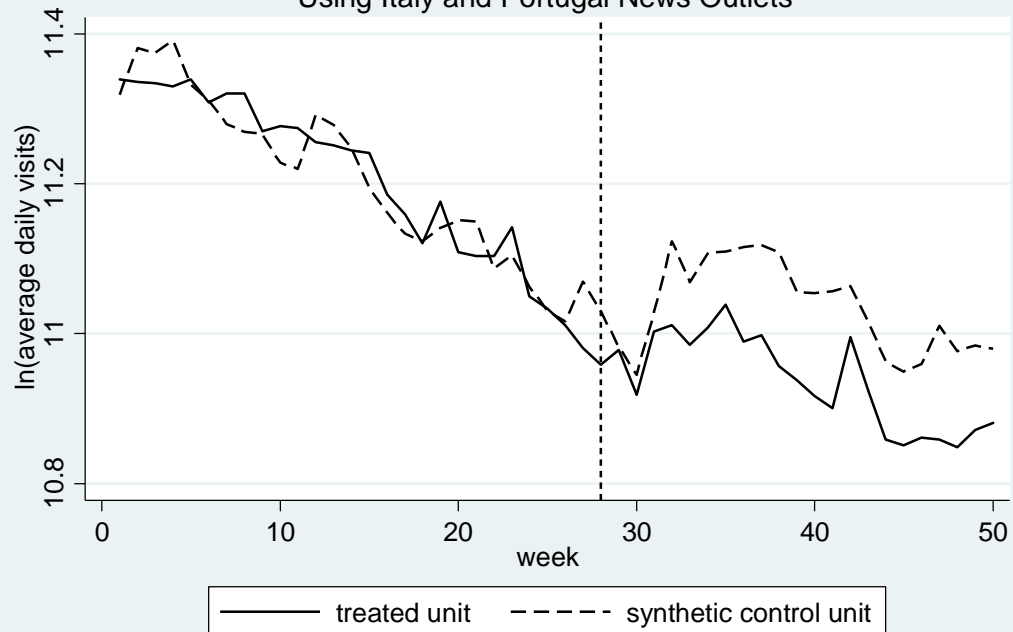


Figure 5C. Synthetic Control Analysis National News Outlets Only
Using France, Germany, Italy and Portugal News Outlets

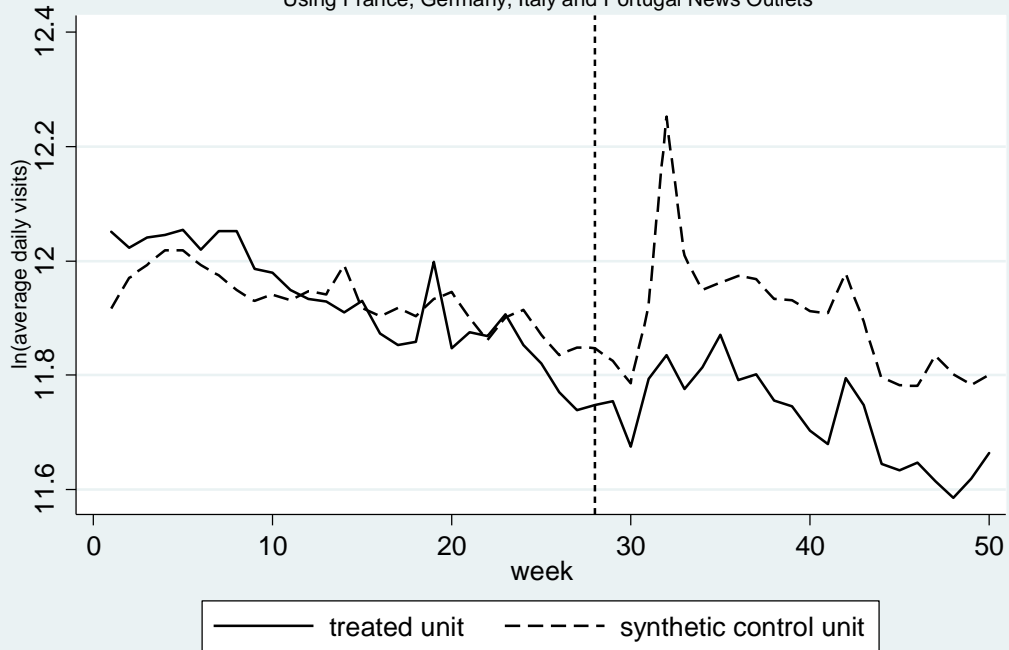


Figure 5D. Synthetic Control Analysis National News Outlets Only
Using Italy and Portugal News Outlets

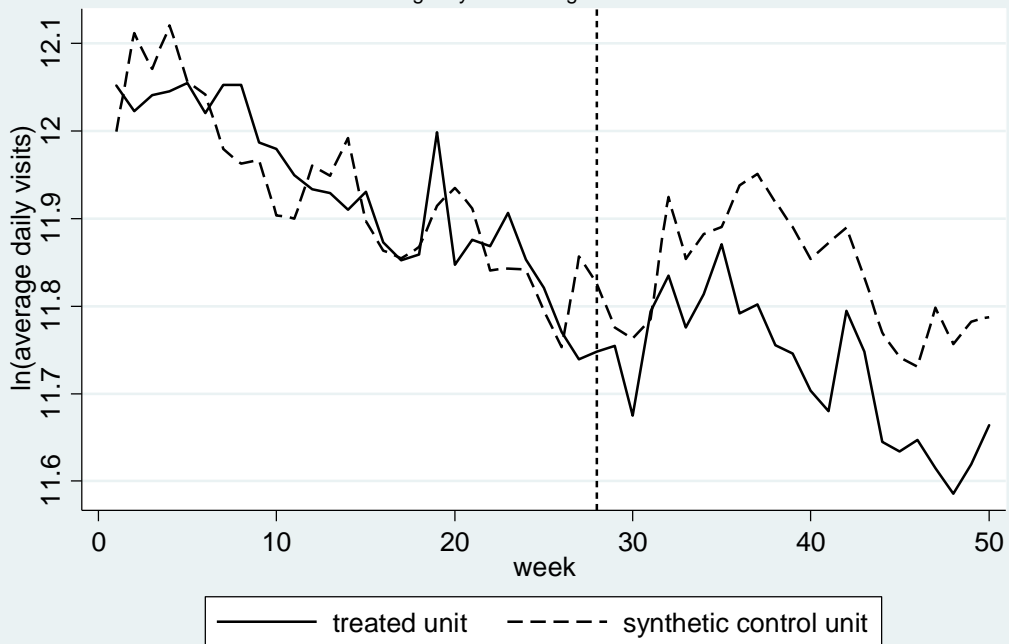


Table 1. Internet User Characteristics in European Union Countries (2014)

	Country Population (Millions)	Household Broadband Access (%)	Internet Users (%)	Users 16-24 (%)	Users 25-34 (%)	Users 35-44 (%)	Users 45-54 (%)	Users 55-74 (%)	Social networks (%)	E-banking (%)	Use E-mail (%)
Spain	45,5	73%	76%	98%	96%	94%	90%	89%	51%	37%	64%
France	65,9	77%	84%	98%	97%	96%	94%	94%	39%	58%	73%
Germany	80,7	87%	86%	99%	99%	97%	95%	92%	42%	49%	80%
Italy	60,7	71%	62%	98%	97%	96%	96%	94%	36%	26%	53%
Portugal	10,4	63%	65%	99%	94%	94%	94%	89%	47%	25%	54%

Data available from Eurostat: <http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/main-tables>.

Table 2. List of Domains per Country

Spain Site	Classif.	France Site	Classif.	Germany Site	Classif.	Italy Site	Classif.
20minutos.es	N, TOP	20minutes.fr	N, TOP	abendblatt.de	N, BOT	adnkronos.com	N, BOT
abc.es	N, TOP	boursorama.com	B, TOP	autobild.de	B, BOT **	agi.it	N, TOP
ara.cat	N, BOT	capital.fr	B, BOT	berliner-zeitung.de	R, BOT **	ansa.it	N, TOP
as.com	S, TOP	challenges.fr	B, BOT	bild.de	N, TOP **	corriere.it	N, TOP
bolsamania.com	B, BOT	eurosport.fr	S, TOP	computerbild.de	B, TOP **	corrieredellosport.it	S, TOP
cincodias.com	B, TOP	footmercato.net	S, BOT	derwesten.de	R, BOT **	diretta.it	S, TOP
diaridegirona.cat	R, BOT	huffingtonpost.fr	N, TOP	deutsche-wirtschafts-n.	B, BOT	ecodibergamo.it	R, BOT
diariocordoba.com	R, BOT	journaldesfemmes	N, TOP	express.de	R, BOT *	fanpage.it	N, TOP
diariodecadiz.es	R, BOT	journaldunet.com	B, TOP	faz.net	N, TOP	gazzetta.it	S, TOP
diariodemallorca.es	R, BOT	ladepeche.fr	R, TOP	finanzen.net	B, TOP **	gds.it	R, BOT
diariodenavarra.es	R, BOT	laprovence.com	R, BOT	focus.de	N, TOP	gelocal.it	R, TOP
diariodesevilla.es	R, BOT	latribune.fr	N, BOT	fr-online.de	R, BOT	huffingtonpost.it	N, TOP
diariosur.es	R, BOT	lavoixdunord.fr	R, BOT	handelsblatt.com	B, BOT *	ilfattoquotidiano.it	N, TOP
diariovasco.com	R, BOT	ledauphine.com	R, BOT	hna.de	R, BOT *	ilgazzettino.it	R, BOT
elcomercio.es	R, BOT	lefigaro.fr	N, TOP	huffingtonpost.de	N, TOP	ilgiornale.it	N, TOP
elconfidencial.com	N, TOP	lemonde.fr	N, TOP	kicker.de	S, TOP	ilgiorno.it	R, BOT
elconfidencialdigital.com	N, BOT	leparisien.fr	R, TOP	ksta.de	R, BOT *	ilmattino.it	R, BOT
elcorreo.com	R, BOT	lepoint.fr	R, TOP	manager-magazin.de	B, BOT	ilmessaggero.it	N, TOP
eldiario.es	N, TOP	leprogres.fr	R, BOT	n24.de	N, TOP **	ilmeteo.it	N, TOP
eldiariomontanes.es	R, BOT	lequipe.fr	R, TOP	news.de	N, BOT	ilpost.it	N, TOP
eleconomista.es	B, TOP	lesechos.fr	N, TOP	rp-online.de	R, BOT *	ilrestodelcarlino.it	N, BOT
elmundo.es	N, TOP	letelegramme.fr	R, BOT	spiegel.de	N, TOP	ilsecoloxix.it	R, BOT
elpais.com	N, TOP	liberation.fr	R, TOP	sport1.de	S, TOP *	ilsole24ore.com	B, TOP
elperiodico.cat	R, BOT	metronews.fr	R, TOP	sportbild.bild.de	S, TOP **	iltempo.it	R, BOT
elperiodico.com	N, TOP	midilibre.fr	R, BOT	stern.de	N, TOP	internazionale.it	N, BOT
elplural.com	N, BOT	ouest-france.fr	R, TOP	sueddeutsche.de	R, TOP	lanazione.it	R, BOT
elpuntavui.cat	R, BOT	rugbyrama.fr	R, BOT	tagesschau.de	N, TOP	lastampa.it	N, TOP
europapress.es	N, TOP	sudouest.fr	R, BOT	tagesspiegel.de	N, BOT	leggo.it	N, BOT
expansion.com	B, TOP	tempsreel.nouvelobs	R, TOP	taz.de	N, BOT	libero.it	N, TOP
heraldo.es	R, BOT			transfermarkt.de	S, TOP **	liberoquotidiano.it	N, TOP
huffingtonpost.es	N, TOP			welt.de	N, TOP **	milannews.it	S, BOT
ideal.es	R, BOT			zeit.de	N, TOP	milanofinanza.it	B, BOT
invertia.com	B, BOT					notizie.it	N, TOP
lainformacion.com	N, BOT					palermotoday.it	R, BOT
laopiniondemalaga.es	R, BOT					panorama.it	N, BOT
larazon.es	N, BOT					quotidiano.net	R, BOT
lasprovincias.es	R, BOT					repubblica.it	N, TOP
lavanguardia.com	N, TOP					romatoday.it	R, BOT
laverdad.es	R, BOT					tuttosport.com	S, TOP
lavozealicia.es	R, TOP					unionesarda.it	R, BOT
levante-emv.com	R, BOT						
libertaddigital.com	N, TOP						
lne.es	R, BOT						
marca.com	S, TOP						
mundodeportivo.com	S, TOP						
naciodigital.cat	R, BOT						
periodistadigital.com	N, BOT						
sport.es	S, TOP						
ultimahora.es	R, BOT						
vilaweb.cat	R, BOT						
vozpopuli.com	N, BOT						

Note: Outlets classification: N= National; R= Regional; B= Business; S= Sports. TOP=Top ranked outlets; BOT=Bottom ranked outlets. In the case of German outlets, * VG Media and ** Axel Springer. Note that Axel Springer outlets also belong to VG Media.

Table 3A. Summary Statistics (Spain, Italy and France)

Variable	Obs	Mean	Std. Dev.	Min	Max
Daily Visits	48,545	245358.70	418584.90	2236.01	6329730
Pages per Visit	47,951	3.72	2.44	1	116.72
Visit Duration	47,951	622.88	465.10	0	7002
Bounce Rate	47,951	0.53	0.11	0	1
Spain	48,545	0.44	0.50	0	1
Italy	48,545	0.34	0.47	0	1
France	48,545	0.22	0.41	0	1
National	48,545	0.38	0.49	0	1
Regional	48,545	0.41	0.49	0	1
Business	48,545	0.10	0.30	0	1
Sports	48,545	0.11	0.31	0	1
National Rank	48,545	519.80	846.17	8	5834
TOP Rank?	48,545	0.45	0.50	0	1
% Domestic Visits	48,545	85.89	9.40	58.4	99.4
TOP Int'l Visits?	48,545	0.39	0.49	0	1

News aggregators sites not considered here.

Table 3B. Summary Statistics (Germany, Italy and France)

Variable	Obs	Mean	Std. Dev.	Min	Max
Daily Visits	38,690	302257.4	462201.8	5000	6329730
Pages per Visit	38,668	3.83	2.89	1	130.97
Visit Duration	38,668	428.43	398.28	0	7002
Bounce Rate	38,668	0.54	0.12	0	1
Germany	38,690	0.30	0.46	0	1
VGM	11,680	0.50	0.50	0	1
Axel Springer	11,680	0.31	0.46	0	1
Italy	38,690	0.42	0.49	0	1
France	38,690	0.28	0.45	0	1
National	38,690	0.42	0.49	0	1
Regional	38,690	0.31	0.46	0	1
Business	38,690	0.13	0.34	0	1
Sports	38,690	0.13	0.34	0	1
National Rank	38,690	338.49	515.43	8	3751
TOP Rank?	38,690	0.53	0.50	0	1
% Domestic Visits	38,325	86.23	7.89	65	99.4
TOP Int'l Visits?	38,325	0.45	0.50	0	1

News aggregators sites not considered here.

Table 4A. Summary Statistics per Country Before and After Google News Shutdown in Spain

Variable	France				Italy				Spain			
	Before Shutdown		After Shutdown		Before Shutdown		After Shutdown		Before Shutdown		After Shutdown	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Daily Visits	298514.30	276473.40	295954.90	298912.50	361935.70	675002.70	266406.80	450748.50	180496.40	308733.70	144280.40	250630.60
Pages per Visit	3.88	1.57	3.68	1.64	4.14	2.43	4.20	4.23	3.40	1.99	3.28	1.49
Visit Duration	605.28	484.30	736.86	571.45	388.07	270.57	498.64	398.68	638.26	356.61	872.81	547.99
Bounce Rate	0.54	0.10	0.55	0.11	0.53	0.14	0.53	0.14	0.53	0.09	0.53	0.10
National	0.34	0.48			0.49	0.50			0.32	0.47		
Regional	0.34	0.48			0.31	0.46			0.53	0.50		
Business	0.17	0.38			0.07	0.25			0.08	0.28		
Sports	0.14	0.34			0.13	0.34			0.07	0.25		
National Rank	222.28	196.22			475.76	725.06			699.64	1056.86		
TOP Rank?	0.59	0.49			0.51	0.50			0.34	0.47		
% Domestic Visits	80.72	8.18			92.34	4.10			83.52	10.15		
TOP Int'l Visits?	0.62	0.49			0.33	0.47			0.32	0.47		

Note: Google News sites are not ranked here, because google.com is #1 everywhere. French and Spanish newspapers qualify as top international visits if % domestic visits is 80% or lower, Italian newspapers if 10% or lower.

Table 4B. Summary Statistics per Country Before, During and After Opt-Out Episode in Germany

Variable	France				Italy				Germany			
	Before & After Opt-Out		During		Before & After Opt-Out		During		Before & After Opt-Out		During	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Daily Visits	297691.50	287539.10	288797.70	270170.90	318587.50	585884.60	316036.40	581798.60	282585.40	384608.20	316283.70	447048.30
Pages per Visit	3.78	1.59	4.13	1.87	4.17	3.41	4.07	1.96	3.42	3.05	3.28	2.20
Visit Duration	663.62	528.66	702.74	556.24	439.13	340.16	420.75	323.11	199.88	76.09	199.75	77.93
Bounce Rate	0.54	0.11	0.55	0.10	0.53	0.14	0.53	0.13	0.54	0.10	0.55	0.11
National	0.34	0.48			0.49	0.50			0.41	0.49		
Regional	0.34	0.48			0.31	0.46			0.28	0.45		
Business	0.17	0.38			0.07	0.25			0.19	0.39		
Sports	0.14	0.34			0.13	0.34			0.13	0.33		
National Rank	222.28	196.22			475.76	725.06			250.78	243.90		
TOP Rank?	0.59	0.49			0.51	0.50			0.50	0.50		
% Domestic Visits	80.72	8.18			92.34	4.10			82.52	5.26		
TOP Int'l Visits?	0.62	0.49			0.33	0.47			0.45	0.50		

Note: Google News sites are not ranked here, because google.com is #1 everywhere. French and Spanish newspapers qualify as top international visits if % domestic visits is 80% or lower, Italian newspapers if 10% or lower.

Table 5A. Diff-in-diff of Shutting Down Google News in Spain after 12/16/2014

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Variable	ln(Daily Visits)				ln(Daily Visits)			
Control Group	French Newspapers				Italian Newspapers			
Shutdown*Spain	-0.146*** (-5.056)				-0.0845*** (-3.174)			
Shutdown*Spain*National		-0.0828** (-2.071)				-0.021 (-0.557)		
Shutdown*Spain*Regional		-0.198*** (-5.150)				-0.137*** (-3.716)		
Shutdown*Spain*Business		-0.030 (-0.572)				0.032 (0.622)		
Shutdown*Spain*Sports		-0.187*** (-4.549)				-0.126*** (-3.176)		
Shutdown*Spain*Top 50%			-0.0765* (-1.965)				-0.015 (-0.405)	
Shutdown*Spain*Bottom 50%			-0.182*** (-5.227)				-0.120*** (-3.655)	
Shutdown*Spain*Top Int 50%				-0.0894** (-2.319)				-0.028 (-0.758)
Shutdown*Spain*Bottom Int 50%				-0.173*** (-4.939)				-0.111*** (-3.361)
Constant	14.62*** (4.936)	14.62*** (4.936)	14.62*** (4.936)	14.62*** (4.936)	96.760 (0.288)	98.190 (0.292)	96.730 (0.288)	96.730 (0.288)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES
Observations	32,120	32,120	32,120	32,120	37,960	37,960	37,960	37,960
R-squared	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

Robust t-statistics in parentheses clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5B. Shutting Down Google News Spain Deleting Four Weeks of Charlie Hebdo Observations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Variable	ln(Daily Visits)				ln(Daily Visits)			
Control Group	French Newspapers				Italian Newspapers			
Shutdown*Spain	-0.0741*** (-2.973)				-0.0798*** (-2.699)			
Shutdown*Spain*National		-0.014 (-0.345)				-0.019 (-0.455)		
Shutdown*Spain*Regional		-0.123*** (-3.469)				-0.129*** (-3.315)		
Shutdown*Spain*Business		0.039 (0.765)				0.033 (0.625)		
Shutdown*Spain*Sports		-0.124*** (-3.020)				-0.130*** (-2.949)		
Shutdown*Spain*Top 50%			-0.006 (-0.152)				-0.012 (-0.278)	
Shutdown*Spain*Bottom 50%			-0.109*** (-3.486)				-0.115*** (-3.272)	
Shutdown*Spain*Top Int 50%				-0.023 (-0.611)				-0.029 (-0.702)
Shutdown*Spain*Bottom Int 50%				-0.0984*** (-3.090)				-0.104*** (-2.926)
Constant	14.42*** (4.879)	14.42*** (4.879)	14.42*** (4.879)	14.42*** (4.879)	61.62 (0.156)	63.05 (0.160)	61.59 (0.156)	61.60 (0.156)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES
Observations	30,184	30,184	30,184	30,184	35,672	35,672	35,672	35,672
R-squared	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

This table repeats the exercise in Table 5 dropping observations from January 7th to February 4th 2015 (both included).

Results are qualitatively the same if dropping from January 7th to January 21st or January 28th 2015.

Robust t-statistics in parentheses clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6. Diff-in-diff by Online Newspaper Type, National Ranking and International Position

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Variable	ln(Daily Visits)							
PANEL A								
Control Group	French Newspapers							
Shutdown*Spain	-0.210*** (-6.423)	-0.179*** (-4.375)	0.045 (0.748)	-0.069 (-1.227)	-0.144*** (-3.899)	-0.141*** (-3.439)	-0.113*** (-3.119)	-0.157*** (-3.855)
Constant	8.976*** (3.163)	1,439*** (4.277)	-555.500 (-0.613)	-3,409** (-2.456)	2,575*** (4.607)	326.5*** (3.213)	-4,704*** (-4.096)	21.55*** (7.461)
Observations	10,585	14,965	2,920	3,650	13,505	18,615	13,505	18,615
R-squared	0.97	0.97	0.97	0.93	0.96	0.94	0.97	0.96
PANEL B								
Control Group	Italian Newspapers							
Shutdown*Spain	-0.0828** (-2.176)	-0.057 (-1.423)	-0.052 (-0.728)	-0.138*** (-4.760)	-0.0636** (-2.138)	-0.102** (-2.462)	-0.100** (-2.53)	-0.076** (-2.26)
Constant	-501.900 (-0.366)	-38.510 (-0.201)	200.800 (0.187)	-466.700 (-0.779)	-159.000 (-0.288)	-169.900 (-0.444)	-970.397* (-2.002)	-243.530 (-1.25)
Observations	14,965	16,425	3,650	2,920	15,695	22,265	12,410	25,550
R-squared	0.95	0.95	0.99	0.97	0.94	0.93	0.98	0.95
Sample	National	Regional	Business	Sports	TOP 50%	BOTTOM 50%	TOP INT 50%	BOTTOM INT 50%

All specifications include site and date fixed effects, as well as group-specific time trends.

Robust t-statistics in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7A. Diff-in-diff of Evolution of Effect by Online Newspaper Type and National Ranking Position

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep Variable	ln(Daily Visits)								
Control Group	French Newspapers								
First Week After*Spain	-0.035 (-1.413)	-0.0683* (-1.813)	-0.031 (-0.740)	0.048 (0.796)	-0.013 (-0.216)	-0.030 (-0.858)	-0.041 (-1.281)	-0.030 (-0.944)	-0.018 (-0.474)
Second Week After*Spain	-0.020 (-0.559)	-0.038 (-0.791)	-0.055 (-1.171)	0.016 (0.143)	0.047 (0.399)	0.025 (0.531)	-0.0759* (-1.765)	0.024 (0.502)	-0.059 (-1.393)
Third Week After*Spain	-0.0609* (-1.728)	-0.063 (-1.558)	-0.087 (-1.453)	-0.105 (-1.000)	0.041 (0.392)	-0.023 (-0.610)	-0.110* (-1.868)	-0.016 (-0.409)	-0.091 (-1.642)
Fourth Week After*Spain	0.220** (2.345)	-0.079 (-1.588)	-0.294*** (-3.581)	0.037 (0.311)	0.354*** (5.013)	-37.26*** (-4.581)	-2.210** (-2.301)	-13.13*** (-4.200)	-0.151* (-1.686)
Fifth Week After*Spain	-0.216*** (-4.998)	-0.348*** (-5.393)	-0.213*** (-3.419)	-0.143*** (-3.083)	0.099 (1.789)	-0.257*** (-4.495)	-0.148*** (-2.781)	-0.208*** (-3.707)	-0.181*** (-3.102)
Sixth Week After*Spain	-0.110*** (-3.564)	-0.154*** (-3.608)	-0.122** (-2.560)	-0.067 (-1.100)	0.021 (0.282)	-0.144*** (-4.450)	-0.078 (-1.565)	-0.079 (-1.647)	-0.101** (-2.297)
Seventh Week + After*Spain	-0.137*** (-4.386)	-0.182*** (-3.366)	-0.153*** (-3.079)	-0.119** (-2.320)	0.055 (0.728)	-0.146*** (-4.741)	-0.132** (-2.266)	-0.136*** (-3.327)	-0.125** (-2.634)
Fourth Week After*France	0.770*** (6.066)	0.697*** (4.857)	0.438*** (4.925)	0.181 (1.209)	0.179** (3.472)	-36.71*** (-4.508)	-1.738* (-1.815)	-12.70*** (-4.093)	0.493*** (4.601)
Constant	14.70*** (4.968)	9.024*** (3.140)	1,230*** (3.363)	-3,766** (-2.570)	45.85*** (5.681)	2,575*** (4.610)	311.0** (2.420)	-4,970*** (-4.198)	21.53*** (7.391)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample	All	National	Regional	Business	Sports	TOP 50%	BOTTOM 50%	TOP INT 50%	BOTTOM INT 50%
Observations	32,120	10,585	14,965	3,650	2,920	13,505	18,615	13,505	18,615
R-squared	0.97	0.97	0.97	0.93	0.97	0.96	0.94	0.97	0.96

Robust t-statistics in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 7B. Diff-in-diff of Evolution of Effect by Online Newspaper Type and National Ranking Position

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep Variable	ln(Daily Visits)								
Control Group	Italian Newspapers								
First Week After*Spain	-0.016 (-0.481)	-0.045 (-0.799)	0.035 (0.832)	-0.013 (-0.153)	-0.018 (-0.558)	-0.029 (-0.534)	-0.002 (-0.0532)	-0.010 (0.370)	-0.018 (-0.38)
Second Week After*Spain	-0.038 (-0.970)	-0.101 (-1.574)	0.024 (0.622)	0.098 (1.442)	-0.006 (-0.0536)	-0.053 (-0.874)	-0.018 (-0.401)	-0.006 (0.140)	-0.053 (-0.98)
Third Week After*Spain	-0.018 (-0.508)	-0.006 (-0.122)	-0.029 (-0.461)	-0.015 (-0.174)	0.000 (0.004)	0.023 (0.580)	-0.058 (-1.057)	-0.041 (0.930)	-0.005 (-0.11)
Fourth Week After*Spain	-0.131** (-2.307)	-0.076 (-0.758)	-0.160** (-2.587)	-0.158** (-3.186)	0.026 (0.283)	-0.018 (-0.209)	-0.210*** (-3.179)	-0.202*** (3.770)	-0.094 (-1.23)
Fifth Week After*Spain	-0.108** (-2.165)	-0.098 (-1.089)	-0.053 (-1.088)	-0.293*** (-6.185)	-0.057 (-0.730)	-0.082 (-1.010)	-0.129** (-2.397)	-0.171*** (3.220)	-0.076 (-1.12)
Sixth Week After*Spain	-0.115*** (-2.938)	-0.073 (-1.192)	-0.096* (-1.825)	-0.294*** (-6.598)	-0.139 (-1.415)	-0.0957* (-1.918)	-0.144** (-2.600)	-0.125* (1.930)	-0.108** (-2.26)
Seventh Week + After*Spain	-0.171*** (-4.671)	-0.161** (-2.640)	-0.140*** (-2.778)	-0.267*** (-5.959)	-0.154* (-1.914)	-0.166*** (-3.045)	-0.182*** (-3.897)	-0.176*** (3.300)	-0.169* (-3.49)
Constant	727.0* (1.922)	-1929 (-1.217)	223.700 (1.319)	-0.582 (-0.0709)	-545.000 (-0.458)	-907.000 (-1.369)	412.600 (1.111)	-420.333 (-0.86)	-549.1** (-2.147)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sample	All	National	Regional	Business	Sports	TOP 50%	BOTTOM 50%	TOP INT 50%	BOTTOM INT 50%
Observations	37,960	14,965	16,425	2,920	3,650	15,695	22,265	12,410	25,550
R-squared	0.97	0.95	0.95	0.97	0.99	0.94	0.93	0.98	0.96

Robust t-statistics in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 8A. Seemingly Unrelated Regressions of Diff-in-diff using Engagement Metrics as Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep Variable	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ppv)	ln(1+vd)	ln(1+br)
Control Group	French Newspapers								
Shutdown*Spain	0.003 (0.355)	-0.0376** (-2.325)	0.00424*** (2.765)						
Shutdown*Spain*Top 50%				0.011 (1.173)	-0.0388** (-2.217)	0.00438*** (2.638)			
Shutdown*Spain*Bottom 50%				-0.001 (-0.100)	-0.0370** (-2.238)	0.00417*** (2.662)			
First Week After*Spain							-0.0746*** (-4.497)	-0.125*** (-3.999)	0.002 (0.516)
Second Week After*Spain							-0.0716*** (-4.292)	-0.110*** (-3.497)	0.00835*** (2.803)
Third Week After*Spain							-0.0701*** (-4.183)	-0.198*** (-6.275)	0.00926*** (3.098)
Fourth Week After*Spain							0.0898*** (4.987)	0.053 (1.566)	-0.00700** (-2.178)
Fifth Week After*Spain							0.0438*** (2.587)	0.017 (0.519)	-0.002 (-0.779)
Sixth Week After*Spain							-0.020 (-1.154)	-0.004 (-0.118)	0.002 (0.817)
Seventh Week + After*Spain							0.0660*** (6.343)	0.0396** (2.027)	0.00826*** (4.449)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548
R-squared	0.614	0.806	0.761	0.614	0.806	0.761	0.616	0.807	0.761
Breusch-Pagan Test	chi2(3) = 10971.572		Pr = 0.0000	chi2(3) = 10973.458		Pr = 0.0000	chi2(3) = 10955.74		Pr = 0.0000

Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 8B. Seemingly Unrelated Regressions of Diff-in-diff using Engagement Metrics as Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep Variable	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ppv)	ln(1+vd)	ln(1+br)
Control Group	Italian Newspapers								
Shutdown*Spain	-0.0491*** (-5.323)	0.027 (1.593)	0.00449*** (2.968)						
Shutdown*Spain*Top 50%				-0.0412*** (-4.036)	0.026 (1.372)	0.00464*** (2.765)			
Shutdown*Spain*Bottom 50%				-0.0531*** (-5.594)	0.028 (1.586)	0.00442*** (2.842)			
First Week After*Spain							-0.0734*** (-4.118)	0.003 (0.082)	0.001 (0.193)
Second Week After*Spain							-0.0961*** (-5.364)	0.049 (1.499)	0.0110*** (3.743)
Third Week After*Spain							-0.0488*** (-2.716)	-0.020 (-0.616)	0.00592** (2.008)
Fourth Week After*Spain							-0.008 (-0.401)	0.0631* (1.787)	0.004 (1.408)
Fifth Week After*Spain							-0.021 (-1.155)	0.034 (1.031)	0.003 (0.932)
Sixth Week After*Spain							-0.0404** (-2.206)	0.129*** (3.849)	0.001 (0.490)
Seventh Week + After*Spain							-0.0439*** (-3.931)	-0.0003 (-0.0125)	0.00450*** (2.456)
Site FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time Trends	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366
R-squared	0.60	0.78	0.77	0.60	0.78	0.77	0.60	0.78	0.77
Breusch-Pagan Test	chi2(3) = 15873.282		Pr = 0.0000	chi2(3) = 15874.718		Pr = 0.0000	chi2(3) = 15879.006		Pr = 0.0000

Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 9. Seemingly Unrelated Regressions of Diff-in-diff using Daily Visits and Engagement Metrics as Dependent Variables and Italian News Outlets as Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Variable	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)
Shutdown*Spain	-0.0895*** (-8.098)	-0.0491*** (-5.323)	0.0269 (1.593)	0.0045*** (2.968)								
Shutdown*Spain*Top 50%					-0.0329*** (-2.690)	-0.0412*** (-4.036)	0.0256 (1.372)	0.0046*** (2.765)				
Shutdown*Spain*Bottom 50%					-0.118*** (-10.38)	-0.0531*** (-5.594)	0.0275 (1.586)	0.0044*** (2.842)				
First Week After*Spain									-0.0202 (-0.945)	-0.0734*** (-4.118)	0.0027 (0.082)	0.0006 (0.193)
Second Week After*Spain									-0.0423** (-1.971)	-0.0961*** (-5.364)	0.0491 (1.499)	0.0110*** (3.743)
Third Week After*Spain									-0.0255 (-1.184)	-0.0488*** (-2.716)	-0.0202 (-0.616)	0.0059** (2.008)
Fourth Week After*Spain									-0.136*** (-5.871)	-0.0078 (-0.401)	0.0631* (1.787)	0.0045 (1.408)
Fifth Week After*Spain									-0.112*** (-5.126)	-0.021 (-1.155)	0.0343 (1.031)	0.0028 (0.932)
Sixth Week After*Spain									-0.115*** (-5.224)	-0.0404** (-2.206)	0.129*** (3.849)	0.0015 (0.490)
Seventh Week + After*Spain									-0.137*** (-10.27)	-0.0439*** (-3.931)	-0.0003 (-0.0125)	0.0045** (2.456)
Observations	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366	37,366
R-squared	0.97	0.60	0.78	0.77	0.97	0.60	0.78	0.77	0.97	0.60	0.78	0.77
Breusch-Pagan Test	chi2(6) = 16669.243 Pr = 0.0000				chi2(6) = 16670.633 Pr = 0.0000				chi2(6) = 16680.946 Pr = 0.0000			

This table reports seemingly unrelated regressions of using difference-in-differences for log dailyvisits, log pages per visit, log visit duration (in seconds), and log of bounce rate. The control group are Italian news outlets. All specifications include group-specific time trends, site and date fixed effects. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 10. Advertising Effects of Google News Shutdown on Online News Outlets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	ln(Daily Revenue)	ln(Ad Intensity)	ln(Revenue/Ad Intensity)	ln(Daily Revenue)	ln(Ad Intensity)	ln(Revenue/Ad Intensity)	ln(Daily Revenue)	ln(Ad Intensity)	ln(Revenue/Ad Intensity)
Treatment Group	Online Newspapers								
Control Group	Non News Reporting Webpages								
Newspaper*Shutdown	-0.106*** (-3.620)	-0.028* (-1.872)	-0.078*** (-3.415)	-0.136** (-2.039)	0.003 (0.342)	-0.140** (-2.240)	-0.084*** (-3.134)	-0.041** (-2.106)	-0.043** (-2.054)
Observations	193,015	193,015	193,015	70,383	70,383	70,383	122,632	122,632	122,632
R-squared	0.787	0.816	0.834	0.777	0.874	0.781	0.748	0.768	0.779
Sample	ALL	ALL	ALL	FIRST PAGE	FIRST PAGE	FIRST PAGE	OTHER PAGES	OTHER PAGES	OTHER PAGES
Site*Advertiser*FrontPage FE	YES	YES	YES	NO	NO	NO	NO	NO	NO
Site*Advertiser FE	NO	NO	NO	YES	YES	YES	YES	YES	YES
Date FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Group-Specific Time Trends	YES	YES	YES	YES	YES	YES	YES	YES	YES

This table shows different set of diff-in-diff results using as dependent variables the log of daily advertising per site and advertiser, the advertising intensity per site and advertiser, and the revenues per advertising intensity per site and advertiser, in the front page and others. Advertising intensity is a measure of number of ads per day, site and advertiser. We compare online newspapers to other webpages non-news related webpages. All specifications control for different time trends across groups. Columns (1) to (3) include site*advertiser*frontpage and date fixed effects. Columns (4) to (9) use site*advertiser fixed effects and constrain the sample to observations for the front page or other pages, respectively. Robust t-statistics in parentheses, clustered at the advertiser level. *** p<0.01, ** p<0.05, * p<0.1.

Table 11. Advertising Effects of Google News Shutdown on Online News Outlets

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	ln(Daily Revenue)	ln(Ad Intensity)	ln(Daily Advertisers)	ln(Revenue/Advertiser)	ln(Ad Intensity/Advertiser)	ln(Revenue/Ad Intensity)
PANEL A						
Treatment Group	Online Newspapers					
Control Group	Non News Reporting Webpages					
Sample	All Observations					
Specification	Site*FrontPage and Date Fixed Effects, plus Group-Specific Time Trends					
Newspaper*Shutdown	-0.356 (-1.396)	-0.014 (-0.264)	-0.124* (-1.737)	-0.319* (-1.719)	0.019 (0.466)	-0.351** (-2.241)
Observations	4,405	4,405	4,405	4,405	4,405	4,405
R-squared	0.87	0.97	0.91	0.85	0.93	0.93
PANEL B						
Treatment Group	Online Newspapers					
Control Group	Non News Reporting Webpages					
Sample	Front Page Observations					
Specification	Site, Week and Day of the Week Fixed Effects, plus Group-Specific Time Trends					
Newspaper*Shutdown	-0.575* (-1.762)	-0.131** (-2.327)	-0.197* (-1.666)	-0.464* (-1.837)	0.008 (0.279)	-0.461* (-1.762)
Observations	2,227	2,227	2,227	2,227	2,227	2,227
R-squared	0.79	0.88	0.88	0.77	0.66	0.80
PANEL C						
Treatment Group	Online Newspapers					
Control Group	Non News Reporting Webpages					
Sample	Other Pages Observations					
Specification	Site, Week and Day of the Week Fixed Effects, plus Group-Specific Time Trends					
Newspaper*Shutdown	-0.012 (-0.0622)	-0.0662*** (-4.649)	0.036 (0.211)	-0.136 (-1.040)	-0.0858* (-1.694)	-0.028 (-0.350)
Observations	2,178	2,178	2,178	2,178	2,178	2,178
R-squared	0.89	0.89	0.91	0.85	0.84	0.91

This table shows two different set of diff-in-diff results using as dependent variables the log of daily advertising revenues per site, the advertising intensity per site, the daily number of advertisers per site, the revenue per advertiser, the advertising intensity per advertiser, and the revenue per advertising intensity, in the front page and other pages separately. Advertising intensity is a measure of number of ads per day in a site per advertiser.

PANEL A compares online newspapers to other webpages non-news related using all data. PANEL B uses data only from front pages, and PANEL C uses data only from other pages.

Robust t-statistics in parentheses, clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.

Table 12. Diff-in-Diffs of Opting-out Period in Germany, October and November 2014 Using Italian Newspapers as Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Variable	ln(Daily Visits)							
Opt Out*Germany	0.008 (0.184)	0.020 (0.418)	0.032 (0.705)	0.033 (0.736)				
Opt Out*VGM		-0.024 (-0.701)			-0.024 (-0.688)			
Opt Out*Axel Springer			-0.0763** (-2.506)			-0.0763** (-2.459)		-0.081* (-1.89)
1st Week Opt Out*Axel Springer				-0.0971** (-2.621)			-0.0814** (-2.074)	
2nd Week Opt Out*Axel Springer				-0.0740* (-1.902)			-0.0899** (-2.177)	
Site FE	YES	YES	YES	YES	YES	YES	YES	NO
Day FE	YES	YES	YES	YES	YES	YES	YES	NO
Trends	YES	YES	YES	YES	YES	YES	YES	YES
Control	Italy	Italy	Italy	Italy	Germany	Germany	Germany	Germany
Observations	28,105	28,105	28,105	28,105	11,680	11,680	11,680	10,220
R-squared	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.01

Robust t-statistics are in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.
Column 8 reports diff-in-diff estimates with propensity score matching using command in stata "diff" with the kernel probit and common support options. The variables used for propensity score are dummies national, regional, business and sports, and variables for national rank and percentage of domestic visits.
Treatment in the propensity score matching in column 8 is whether the site belongs to Axel Springer.

Table 13. Diff-in-Diffs Germany Opt-out Robustness Check

	(1)	(2)	(3)	(4)
Dep Variable	ln(Daily Visits)			
Opt Out*Axel Springer	-0.0645* (-1.964)		-0.0830*** (-2.934)	
1st Week Opt Out*Axel Springer		-0.070 (-1.528)		-0.0881** (-2.495)
2nd Week Opt Out*Axel Springer		-0.0784** (-2.117)		-0.0966** (-2.388)
Opt Out*TOP 50%			0.0457 (1.329)	0.046 (1.335)
Opt Out*Regional	0.0466 (1.432)	0.0466 (1.432)		
Opt Out*Sports	0.043 (0.657)	0.0441 (0.680)		
Opt Out*Business	-0.0675 (-1.354)	-0.0664 (-1.322)		
Constant	11.08*** (176.200)	11.08*** (176.200)	11.08*** (174.800)	11.08*** (174.800)
Site FE	YES	YES	YES	YES
Day FE	YES	YES	YES	YES
Trends	YES	YES	YES	YES
Observations	11,680	11,680	11,680	11,680
R-squared	0.95	0.95	0.95	0.95

Robust t-statistics are in parentheses and clustered at the site level.

*** p<0.01, ** p<0.05, * p<0.1

Appendix A: Internet Advertising Market in Spain

The internet advertising market in Spain underwent an important transformation during the time period of our study not only due to the use of new information management technologies, but also due to the appearance of several new agents that facilitated the coordination of publishers and advertisers. Indeed, in recent years direct contracting to sell ad inventory has in part been replaced by programmatic advertisement. With direct contracting, advertisers (or their agencies) request advertisement space from news outlets for their marketing campaigns. Several agents can intervene in this process. Ad Networks buy unsold inventory from publishers that they later resell through the use of technologies that help categorizing, packing and selling such slot inventory. Ad Exchanges facilitate the commercialization of ad inventory from multiple ad networks. Instead of negotiating prices, buyers and sellers of slots may predetermine prices and audience characteristics they are interested in (many consider ad exchanges as the origin of programmatic advertisement). Additionally, both publishers and advertisers may use Ad Servers, which are agents that use specialized software to provide slots, count them, maximize revenues, and monitor the progress of an advertising campaign.

An important drawback of direct advertisement is that it implies costly negotiations, and as a result news outlets and advertisers usually negotiate their campaigns in an annual base only. Otherwise, domains sell slots to advertisers contextually targeting users “in advance.” In such case, domains take into account the knowledge that the advertiser has about the outlet's audience.

Since 2008, programmatic advertising has drastically improved the management of ad inventories. Programmatic advertising is now able to efficiently manage buying and selling ads at large scale in real time, through the use of management algorithms that maximize the impact of each impression. We may classify programmatic advertising in two main types: (1) Real-Time Bidding (RTB) is an auction-based mechanism by which ad inventory is bought and sold on a per-impression basis, via programmatic instantaneous auction. Buyers bid on an impression and, if they win, their ad is instantly displayed on the publisher's site. Ad exchanges are auction-based marketplaces that facilitate this process. (2) Programmatic Direct is a process that permits advertisers to reserve an ad slot they will use in the future while fixing in advance the price and the characteristics of the targeted audience of their advertisement campaign. While programmatic direct is similar to traditional contracting mechanisms, they differ in that programmatic direct uses programmatic technology to automatize and simplify the transactions. In addition, it allows advertisers to individually target users taking into account their individual preferences and interests through previously gathered data.

Because direct contracting is associated with higher returns per ad slot, news outlets prioritize their use. The remaining unsold ad inventory is then offered through programmatic advertisement. Even though news outlets benefit from programmatic advertisement in that they are now able to monetize all their inventory, they have little market power and the prices they can charge through programmatic advertisement are usually low and determined

by market forces. However, in recent years large brands that target their marketing campaigns toward national and international level have transitioned from direct to programmatic advertisement because this mechanism improves the coordination of their campaigns and reduces their total costs.

The final prices paid for insertions depend on factors such as the location of the ad in the webpage, exposure time, or the impact of the advertisement campaign. The prices also depend on the type of ad, which differ between standardized banners, pop-ups and videos. There are also several ways to contract for ads: (1) cost per thousand impressions, CPM, (2) cost per click, CPC; (3) cost per contact person (used to create a database of potential consumers), CPL; or (4) cost per action, which can be one click, a subscription or a sale, CPA. In general, large news outlets want to guarantee a minimum revenue for their inventory, and as a result CPM has become the preferred pricing model.

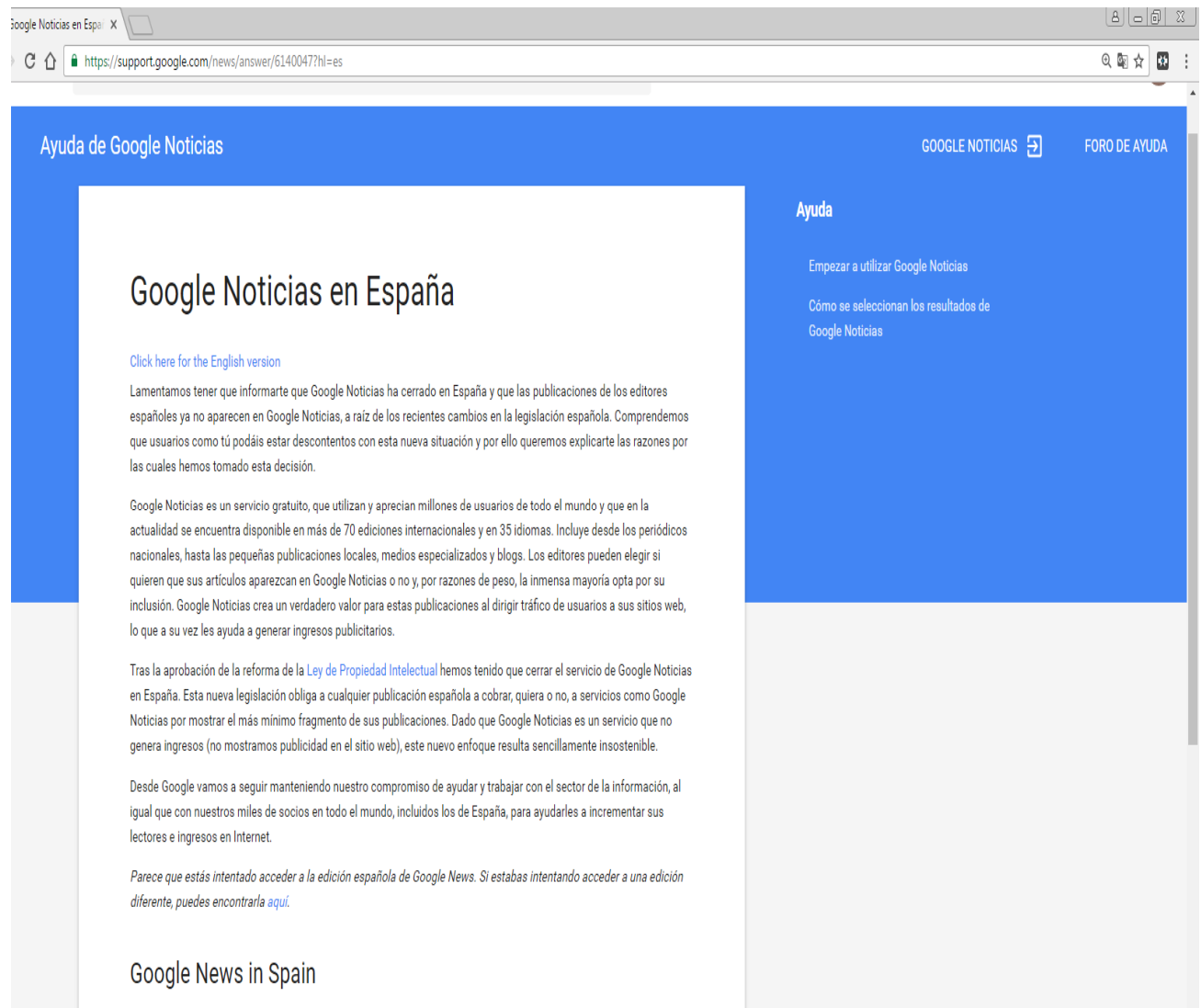
By 2015, programmatic transactions for online display advertising accounted for a significant share in several European countries. In the United Kingdom, it represented 47% of online display advertising spending, 37% in France, and 30% in Italy.⁴⁹ According to IAB (2014), programmatic advertisement represented about 20% of the total spending on digital publicity in Spain.⁵⁰

⁴⁹ Grece, C. (2016). The online advertising market in the EU -- Update 2015 and Focus on programmatic advertising, European Audiovisual Observatory, Strasbourg.

⁵⁰ IAB (2014) El Libro Blanco de la Compra Programática. See also at <https://iabspain.es/wp-content/uploads/Libro-blanco-de-Compra-Programatica-y-RTB.pdf>.

APPENDIX B

Figure A1. Webpage Google News Spain Post-Shutdown



APPENDIX B

Table A1. List of Portuguese Sites

Domain	Classification
abola.pt	S, TOP
record.pt	S, TOP
publico.pt	N, TOP
jn.pt	N, TOP
observador.pt	N, TOP
cmjornal.pt	N, TOP
dn.pt	N, TOP
meusresultados.com	S, TOP
ojogo.pt	S, TOP
noticiasaoiminuto.com	N, TOP
zerozero.pt	S, TOP
jornaldenegocios.pt	B, TOP
dinheirovivo.pt	B, TOP
aeiou.pt	N, TOP
sabado.pt	N, BOT
dnoticias.pt	R, BOT
acorianooriental.pt	R, BOT
jornaldeleiria.pt	R, BOT
omirante.pt	R, BOT
diariocoimbra.pt	R, BOT
funchalnoticias.net	R, BOT
asbeiras.pt	R, BOT
lusa.pt	N, BOT
diarioaveiro.pt	R, BOT

Outlets classification: N= National;
R= Regional; B= Business; S= Sports.
TOP=Top ranked outlets;
BOT=Bottom ranked outlets.

Table A2. List of Domains ARCE

News Outlets	Non-News Outlets
20minutos.es *	atrapalo.com
abc.es *	autocasion.com
ara.cat	bebesymas.com
as.com *	coches.net
bolsamania.com	ebay.es
cincodias.com *	edreams.es
diaridegirona.cat	emprendedores.es
diariocordoba.com	enalquiler.com
diariodecadiz.es	enfemenino.com
diariodemallorca.es	forocoches.com
diariodenavarra.es	fotogramas.es
diariodesevilla.es	guiadelocio.com
diariosur.es	guiainfantil.com
diariovasco.com	idealista.com
elcomercio.es	infojobs.net
elconfidencial.com *	mibebeyyo.com
elcorreo.com *	minijuegos.com
eldiario.es	muycomputer.com
eldiariomontanes.es	muyinteresante.es
eleconomista.es *	paginasamarillas.es
elmundo.es *	pisos.com
elpais.com *	rincondelvago.com
elperiodico.com *	rumbo.es
europapress.es	sem.educaweb.com
expansion.com *	telva.com
heraldo.es	ticketmaster.es
ideal.es	traveler.es
invertia.com	universia.es
lainformacion.com	viamichelin.es
laopiniondemalaga.es	vibbo.com
larazon.es	xataka.com
lasprovincias.es	
lavanguardia.com *	
laverdad.es	
lavozdegallicia.es *	
lavozdigital.es	
levante-emv.com	
libertaddigital.com	
lne.es	
marca.com *	
mitele.es	
mundodeportivo.com *	
periodistadigital.com	
publico.es *	
sport.es *	
ultimahora.es	
vilaweb.cat	

*News sites with above median advertising revenues.

Table A3. Seemingly Unrelated Regressions of Diff-in-diff using Daily Visits and Engagement Metrics as Dependent Variables and French News Outlets as Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Variable	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)	ln(1+ldvisits)	ln(1+ppv)	ln(1+vd)	ln(1+br)
Shutdown*Spain	-0.151*** (-14.14)	-0.0178** (-2.032)	-0.0651*** (-3.953)	0.0014 (0.889)								
Shutdown*Spain*Top 50%					-0.0461*** (-4.050)	0.0109 (1.173)	-0.0388** (-2.217)	0.0044*** (2.638)				
Shutdown*Spain*Bottom 50%					-0.131*** (-12.21)	-0.0009 (-0.100)	-0.0370** (-2.238)	0.0042*** (2.662)				
First Week After*Spain									-0.0428** (-2.126)	-0.0746*** (-4.497)	-0.125*** (-3.999)	0.0015 (0.516)
Second Week After*Spain									-0.0236 (-1.165)	-0.0716*** (-4.292)	-0.110*** (-3.497)	0.0084*** (2.803)
Third Week After*Spain									-0.0635*** (-3.125)	-0.0701*** (-4.183)	-0.198*** (-6.275)	0.0093*** (3.098)
Fourth Week After*Spain									-0.545*** (-24.94)	0.0898*** (4.987)	0.0530 (1.566)	-0.007** (-2.178)
Fifth Week After*Spain									-0.205*** (-9.980)	0.0438*** (2.587)	0.0165 (0.519)	-0.0024 (-0.779)
Sixth Week After*Spain									-0.0903*** (-4.361)	-0.0197 (-1.154)	-0.0038 (-0.118)	0.0025 (0.817)
Seventh Week + After*Spain									-0.0039 (-0.309)	0.0660*** (6.343)	0.0396** (2.027)	0.0083*** (4.449)
Observations	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548	31,548
R-squared	0.97	0.62	0.81	0.76	0.97	0.62	0.81	0.76	0.97	0.61	0.81	0.76
Breusch-Pagan Test	chi2(6) = 11117.820 Pr = 0.0000				chi2(6) = 10985.192 Pr = 0.0000				chi2(6) = 10971.681 Pr = 0.0000			

This table reports seemingly unrelated regressions of using difference-in-differences for log dailyvisits, log pages per visit, log visit duration (in seconds), and log of bounce rate. The control group are French news outlets. All specifications include group-specific time trends, site and date fixed effects. Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A4. SUREG of Advertising Effects of Google News Shutdown on Online News Outlets

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	ln(Ad Intensity)	ln(Revenue/ Ad Intensity)	ln(Ad Intensity)	ln(Revenue/ Ad Intensity)	ln(Ad Intensity)	ln(Revenue/ Ad Intensity)
Treatment Group	Online Newspapers					
Control Group	Non News Reporting Webpages					
Newspaper*Shutdown	-0.0227*** (-2.945)	-0.0687*** (-4.138)	-0.00841 (-0.412)	-0.0946 (-1.611)	-0.0263* (-1.924)	-0.026 (-1.119)
Observations	193,015	193,015	70,383	70,383	122,632	122,632
R-squared	0.01	0.01	0.01	0.01	0.01	0.02
Sample	All	All	Front Page	Front Page	Other Pages	Other Pages
Breusch-Pagan Test	chi2(1) = 0.295 Pr = 0.5867		chi2(1) = 6268.591 Pr = 0.0000		chi2(1) = 11367.872 Pr = 0.0000	

All specifications in this table include site, week and week-day fixed effects, plus group-specific time trends. The observational unit is at the site-advertiser-page type-day level. The dependent variables (log of advertising intensity and log of revenue per ad intensity) are demeaned at. z-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A5. SUREG of Advertising Effects of Google News Shutdown

	(1)	(2)	(3)	(4)
Dep. Var.	ln(Ad Intensity)	ln(Daily Advertisers)	ln(Ad Intensity/ Advertiser)	ln(Revenue/ Ad Intensity)
PANEL A				
Treatment Group	Online Newspapers			
Control Group	Non News Reporting Webpages			
Sample	All Observations (Site/Date level)			
Specification	Site, Week and Week Day FE, plus Time Trends			
Newspaper*Shutdown	-0.056 (-1.387)	-0.316*** (-9.239)	0.248*** (7.161)	-0.403*** (-3.738)
Observations	2,342	2,342	2,342	2,342
R-squared	0.90	0.86	0.87	0.86
Breusch-Pagan Test	chi2(6) = 2400.586 Pr = 0.0000			
PANEL B				
Treatment Group	Online Newspapers			
Control Group	Non News Reporting Webpages			
Sample	All Observations (Site/Page Type/Date level)			
Specification	Site*FrontPage, Week and Week Day FE, plus Time Trends			
Newspaper*Shutdown	-0.029 (-0.631)	-0.132*** (-2.654)	0.015 (0.551)	-0.349*** (-4.773)
Observations	4,405	4,405	4,405	4,405
R-squared	0.96	0.88	0.93	0.93
Breusch-Pagan Test	chi2(6) = 2912.580 Pr = 0.0000			
PANEL C				
Treatment Group	Online Newspapers			
Control Group	Non News Reporting Webpages			
Sample	Front Page Observations (Site/Date level)			
Specification	Site, Week and Week Day FE, plus Time Trends			
Newspaper*Shutdown	0.012 (0.198)	-0.208*** (-2.804)	0.0401* (1.906)	-0.619*** (-4.681)
Observations	2,227	2,227	2,227	2,227
R-squared	0.88	0.88	0.66	0.80
Breusch-Pagan Test	chi2(6) = 1615.164 Pr = 0.0000			
PANEL D				
Treatment Group	Online Newspapers			
Control Group	Non News Reporting Webpages			
Sample	Other Pages Observations (Site/Date level)			
Specification	Site, Week and Week Day FE, plus Time Trends			
Newspaper*Post 12/16/14	-0.022 (-0.603)	-0.030 (-0.536)	0.004 (0.081)	-0.079 (-1.327)
Observations	2,178	2,178	2,178	2,178
R-squared	0.89	0.86	0.84	0.91
Breusch-Pagan Test	chi2(6) = 1745.628 Pr = 0.0000			

z-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A6. Diff-in-Diffs of Opting-out Period in Germany, October and November 2014

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Variable	ln(Daily Visits)							
Opt Out*Germany	0.071*** (3.04)	0.083*** (2.72)	0.095*** (3.62)	0.096*** (3.72)				
Opt Out*VGM		-0.023 (-.70)			-0.024 (-0.688)			
Opt Out*Axel Springer			-0.076** (-2.50)			-0.0763** (-2.459)		-0.081* (-1.89)
1st Week Opt Out*Axel Springer				-0.087** (-2.32)			-0.0814** (-2.074)	
2nd Week Opt Out*Axel Springer				-0.087** (-2.22)			-0.0899** (-2.177)	
Site FE	YES	YES	YES	YES	YES	YES	YES	NO
Day FE	YES	YES	YES	YES	YES	YES	YES	NO
Trends	YES	YES	YES	YES	YES	YES	YES	YES
Control	France	France	France	France	Germany	Germany	Germany	Germany
Observations	22,265	22,265	22,265	22,265	11,680	11,680	11,680	10,220
R-squared	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.01

Robust t-statistics are in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.
Column 8 reports diff-in-diff estimates with propensity score matching using command in stata "diff" with the kernel probit and common support options. The variables used for propensity score are dummies national, regional, business and sports, and variables for national rank and percentage of domestic visits. Treatment in the propensity score matching in column 8 is whether the site belongs to Axel Springer.

Table A7. Diff-in-Diffs using Engagement Metrics Germany Opt-out

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Variable	ln(1+pages per visit)		ln(1+visit duration)		ln(1+bounce rate)	
Opt Out*Axel Springer	0.0163 (0.938)		-0.0333 (-0.984)		0.00161 (0.347)	
1st Week Opt Out*Axel Springer		-0.00586 (-0.231)		-0.0523 (-1.510)		0.00489 (0.986)
2nd Week Opt Out*Axel Springer		0.033 (1.374)		-0.0229 (-0.572)		-0.00106 (-0.189)
Constant	1.110*** (37.130)	1.110*** (37.140)	4.846*** (159.400)	4.846*** (159.400)	0.519*** (87.660)	0.519*** (87.640)
Site FE	YES	YES	YES	YES	YES	YES
Day FE	YES	YES	YES	YES	YES	YES
Trends	YES	YES	YES	YES	YES	YES
Observations	11,680	11,680	11,680	11,680	11,680	11,680
R-squared	0.74	0.74	0.79	0.79	0.87	0.87

Robust t-statistics are in parentheses and clustered at the site level. *** p<0.01, ** p<0.05, * p<0.1.