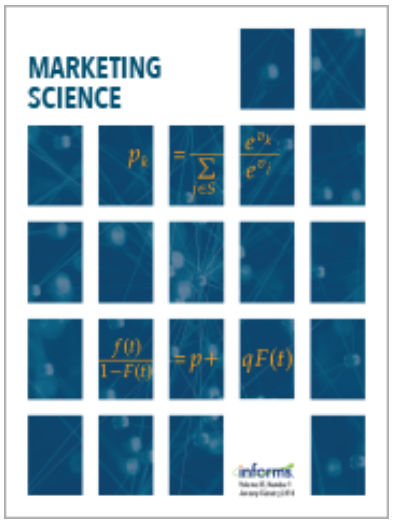


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

What Do News Aggregators Do? Evidence from Google News in Spain and Germany

Joan Calzada, Ricard Gil

To cite this article:
Joan Calzada, Ricard Gil (2020) What Do News Aggregators Do? Evidence from Google News in Spain and Germany. Marketing Science 39(1):134-167. https://doi.org/10.1287/mksc.2019.1150

http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 39, No. 1, January–February 2020, pp. 134–167
ISSN 0732-2399 (print), ISSN 1526-548X (online)

What Do News Aggregators Do? Evidence from Google News in Spain and Germany

Joan Calzada,[a] Ricard Gil[b]

[a] Department of Economics and BEAT, Universitat de Barcelona, 08034 Barcelona, Spain; [b] Smith School of Business, Queen's University, Kingston, Ontario K7L 2P3, Canada

Contact: calzada@ub.edu, http://orcid.org/0000-0002-2976-1677 (JC); ricard.gil@queensu.ca, http://orcid.org/0000-0003-3627-0988 (RG)

Received: June 20, 2017
Revised: April 29, 2018; November 8, 2018; December 10, 2018
Accepted: December 11, 2018
Published Online in Articles in Advance: December 30, 2019

https://doi.org/10.1287/mksc.2019.1150

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 that 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 fewer 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 show 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.

History: Avi Goldfarb served as the senior editor and Anja Lambrecht served as associate editor for this article. This paper has been accepted for the *Marketing Science* Special Issue on Consumer Protection.
Funding: J. Calzada acknowledges financial support from the Spanish Ministerio de Economía y Competitividad [Grant ECO2015-69193-R].
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2019.1150.

Keywords: news aggregators • Google News • market expansion effect • substitution effect • daily visits • engagement metrics

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, accommodations, 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, their advertisement revenues. Although 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 it 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 understanding 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 tradeoff in the various effects of news aggregators on news outlets' audiences (Calzada and Ordoñez 2012,

George and Hogendorn 2012, Dellarocas et al. 2013, 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 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 quasinatural experiment to examine the effect of news aggregators on the number of visits received by news outlets and 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 to the news stories of news outlets if using excerpts of less than seven 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 because of drastic traffic losses, and accepted Google's opt-in conditions.

Our study draws from a rich data set obtained from SimilarWeb containing information on 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 analyses 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 lowest-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 that the impact of the shutdown depends on the outlets' specialization. Specifically, we find that 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 that the shutdown of Google News reduced the average duration and the number of pages per visit while increasing bouncing rates. These findings suggest that the shutdown significantly changed the composition of the consumers who 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. Our

empirical analysis compares advertising revenues and advertising intensity at the site-advertiser-date level across news and nonnews 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 the 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 do we find an 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 who exhibited similar consumption patterns in the past. They find that 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% (bypassing Google News altogether), and clicks on local outlets from the Google News page increased by 13%. In a related paper, George and Hogendorn (2014) use a major redesign in the U.S. edition of Google News on June 30, 2010. Similar to the previous case, the redesign placed a permanent strip of geotargeted local news headlines onto the Google News front page. Using a sample of news visits by U.S. households before and after the introduction of the geotargeted links, the authors find that local news visits increased by less than 1% and that 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 geotargeted 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 that users navigated immediately after visiting Google News or Yahoo! News, they find that Google News users were less likely to visit other news websites after visiting Google News after 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 allow us to test the predictions from the theoretical literature and Section 3. 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 our work is that by Athey et al. (2017b), which independently studies 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 our results (both papers find an overall 10% reduction in the number of daily visits to news outlets), although the data and methodology differ across papers. On the one hand, our data are at the domain level, including information from all visits from desktop users, and their data are at the individual level for only users of Microsoft products, accounting for one-half of personal computer news browsing activity. On the other hand, our methodology compares a sample of Spanish news outlets with a control group of French and Italian news outlets, and

their study uses individual-level browsing data to compare a designated group of Google News users with 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 for 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. (2018) 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, Cagé et al. (2015) examine 84 general information media outlets in France (including newspapers, television channels, radio stations, and news agencies) and track every article that these sites offered online in 2013 with the help of a plagiarism-detection algorithm that quantifies the copy rate between an article and all of the articles previously published about the event. They find that one-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 Google's ban from sponsored search listings in 2010 of 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, although aggregators rely on algorithms, online social network rely on friends to filter and recommend content. Second, because the scope of activities performed in online social networks is far richer, complementarities between news stories and recommending parties may be an important driver of 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. Although 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 then describes the creation of a link fee in Spain and Germany, which motivated 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 countries[8]:

> 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 allows for better targeting of consumers.[9]

More generally, an overall revision of Google News' disputes in Europe shows that 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 a strategy is the launch by Google in April 2015 of the Digital News Initiative (DNI), collaborating with *Les Échos* (France), *FAZ* and *Die Zeit* (Germany), the *Financial Times* and *The Guardian* (the United Kingdom), NRC Group (Netherlands), *La Stampa* (Italy), and *El País* and Grupo Godó (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.[11] The new law established that online outlets posting links and excerpts of news articles that originate elsewhere must pay a link fee (canon) to the original publishers. The creation of the link fee was initially promoted by the publishers association Asociación de Editores de Diarios Españoles (AEDE), which lobbied the government to force news aggregators to compensate them for the use of their content.[12]

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 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 the same from other, yet smaller, Spanish news aggregators, such as Planeta Ludico, NiagaRank, Multifriki, InfoAliment, and Beeeinfo. 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 larger reductions for smaller newspapers (NERA 2015, 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 that 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 that 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 that publishers can prohibit aggregators from using their news articles beyond headlines and short excerpts and that they can charge aggregators a link fee if the aggregators make a larger use of their contents. The main 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 that (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 television 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 that 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 AGOF rankings (https://www.agof.de/en/agof/), and computerbild.de lost its top 10 rank of all AGOF offerings in Germany.[20] Shortly after, on 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 for 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 a departure point that consumers visit a limited number of news outlets, which typically differ on the topics that 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 increase the total amount of reading time instead of doing other leisure activities (Dellarocas et al. 2013). In this context, the different treatment that 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 reduces the number of direct visits to indexed news outlets. A substitution effect occurs when the consumers who visit news aggregators may not visit the news outlets' front pages or click through to their content pages (Dellarocas et al. 2016). Third, 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 are aggregated at the domain-day level, we cannot investigate the impact of the shutdown of Google News in Spain on consumer behavior as Athey et al. (2017b) 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.

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 read more news stories (Jeon 2018). Following this logic, we predict that the lower the brand awareness of an outlet, the

larger the market expansion effect that 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 because of 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 that 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 large brand awareness and brand loyalty. For this type of outlet, news aggregator will have little effect on the number of daily visits, because most readers already know about their existence (no large market expansion effect). Note also that, with multihoming, 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 because of the readers' ideological preferences or because they are niche outlets with a 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 in a particular niche with poor substitutes.

In a second hypothetical scenario, imagine an outlet with low brand awareness and 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 who otherwise would not be aware of its presence (market expansion effect). In contrast, the outlet would not lose visitors because of 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. Although regional outlets may attract visitors from other regions, sports outlets may receive visits from multihoming 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, although they have a strong loyal audience, they also receive many visits through news aggregators from consumers based in other regions.[22]

A third hypothetical scenario is one where an outlet has low brand loyalty combined with low or 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 as 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]

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

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 email (%)
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

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

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 that 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 one-half of a year before and after Google News' shutdown in Spain. Google's removal 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. Discarding Germany as a control group was a no brainer because of its own Google News-"related turbulence" in October 2014. Portugal, however, is very different from Spain in terms of its population size, broadband usage, internet usage, and other demographic and consumption characteristics of internet consumers.[24] Table 1 shows that Spain's internet penetration and household broadband access rates are far higher than in 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 seems ranked in the middle according to the use of e-banking and emails among countries in Table 1, and it only exceeded 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 by using French and Italian news outlets as comparison groups to German news outlets for similar reasons. Because only VG Media sites in Germany opted out from Google News, our analysis takes us in the end to compare VG Media and Axel Springer outlets (a publisher within VG Media) with 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 web pages from television and radio stations and other potential news aggregators, such as MSN or Yahoo. 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; Yahoo! News in Spain (es.noticas.yahoo.com); and 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 that they receive from other countries. Top international outlets are those that receive more than 25% of 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

Table 2. List of Domains per Country

Spain site	Classification	France site	Classification	Germany site	Classification	Italy site	Classification
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[a]	agi.it	N, TOP
ara.cat	N, BOT	capital.fr	B, BOT	berliner-zeitung.de	R, BOT[a]	ansa.it	N, TOP
as.com	S, TOP	challenges.fr	B, BOT	bild.de	N, TOP[a]	corriere.it	N, TOP
bolsamania.com	B, BOT	eurosport.fr	S, TOP	computerbild.de	B, TOP[a]	corrieredellosport.it	S, TOP
cincodias.com	B, TOP	footmercato.net	S, BOT	derwesten.de	R, BOT[a]	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, BO[b]	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[a]	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[b]	ilfattoquotidiano.it	N, TOP
diariovasco.com	R, BOT	ledauphine.com	R, BOT	hna.de	R, BOT[b]	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[b]	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[a]	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[b]	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[b]	ilsole24ore.com	B, TOP
elperiodico.cat	R, BOT	metronews.fr	R, TOP	sportbild.bild.de	S, TOP[a]	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	nouvelobs.com	R, TOP	taz.de	N, BOT	libero.it	N, TOP
heraldo.es	R, BOT			transfermarkt.de	S, TOP[a]	liberoquotidiano.it	N, TOP
huffingtonpost.es	N, TOP			welt.de	N, TOP[a]	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
lavozdegalicia.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						

Notes. In the case of German outlets, VG Media and Axel Springer are noted; however, Axel Springer outlets belong to VG Media. Outlets classification are as follows: BOT, bottom-ranked outlets; N, national; R, regional; B, business; S, sports; TOP, top-ranked outlets.

[a]Axel Springer.

[b]VG Media.

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 it 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 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. However, during the duration of the visit, all of the activities, such as clicking on articles and images, are considered page views. *Pages per visit* indicates the daily page views divided

Table 3. Summary Statistics

Variable	Observations	Mean	Standard deviation	Min	Max
Spain, Italy, and France					
Daily visits	48,545	245,358.70	418,584.90	2,236.01	6,329,730
Pages per visit	47,951	3.72	2.44	1	116.72
Visit duration	47,951	622.88	465.10	0	7,002
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	5,834
Top rank?	48,545	0.45	0.50	0	1
% Domestic visits	48,545	85.89	9.40	58.4	99.4
Top international visits?	48,545	0.39	0.49	0	1
Germany, Italy, and France					
Daily visits	38,690	302,257.4	462,201.8	5,000	6,329,730
Pages per visit	38,668	3.83	2.89	1	130.97
Visit duration	38,668	428.43	398.28	0	7,002
Bounce rate	38,668	0.54	0.12	0	1
Germany	38,690	0.30	0.46	0	1
VG Media	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	3,751
Top rank?	38,690	0.53	0.50	0	1
% Domestic visits	38,325	86.23	7.89	65	99.4
Top international visits?	38,325	0.45	0.50	0	1

Note. News aggregators sites are not considered here.

by the daily visits of the domain. Finally, *Bounce rate* is a 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.

Table 3 shows the summary statistics for all of the variables obtained from SimilarWeb for our Spanish and German analyses, respectively. Table 3 shows 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, which last 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 seven minutes (428.4 seconds).

Tables 4 and 5 separate the data by country as well as 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 of 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 news 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 that 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 4 that all countries experience a decrease in daily visits after the shutdown

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

	France				Italy				Spain			
	Before shutdown		After shutdown		Before shutdown		After shutdown		Before shutdown		After shutdown	
Variable	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Daily visits	298,514.30	276,473.40	295,954.90	298,912.50	361,935.70	675,002.70	266,406.80	450,748.50	180,496.40	308,733.70	144,280.40	250,630.60
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	1,056.86		
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 international visits?	0.62	0.49			0.33	0.47			0.32	0.47		

Notes. Google News sites are not ranked here, because google.com is number one everywhere. French and Spanish newspapers qualify as top international visits if percentage of domestic visits is 80% or lower, and Italian newspapers qualify as top international visits if percentage of domestic visits is 10% or lower. Std. dev., standard deviation.

of Google News in Spain and that the decreases in Spain and Italy are proportionally larger than the decrease in France. Interestingly, Table 5 shows that visits to 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%; and 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 substitution effects as well as 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:

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

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) and $SPAIN_i$ is a dummy variable that takes the value of one if site i belongs to Spain and zero otherwise. $Shutdown_t$ is another dummy variable that takes the value of one if day t is after December 16, 2014 and zero otherwise. We also use country-site–specific fixed effects γ_i and date fixed effects δ_t, and we 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 identically independently and normally distributed (niid) as usual. The main objective of our analysis is to identify the impact of the shutdown on 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 numbers of pages per

Table 5. Summary Statistics per Country Before, During, and After the Opt-Out Episode in Germany

	France				Italy				Germany			
	Before and after opt-out		During		Before and after opt-out		During		Before and after opt-out		During	
Variable	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Daily visits	297,691.50	287,539.10	288,797.70	270,170.90	318,587.50	585,884.60	316,036.40	581,798.60	282,585.40	384,608.20	316,283.70	447,048.30
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 international visits?	0.62	0.49			0.33	0.47			0.45	0.50		

Notes. Google News sites are not ranked here, because google.com is number one everywhere. French and German newspapers qualify as top international visits if percentage of domestic visits is 80% or lower, and Italian newspapers qualify as top international visits if percentage of domestic visits is 10% or lower. Std. dev., standard deviation.

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 difference-in-differences 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. Although 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:

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

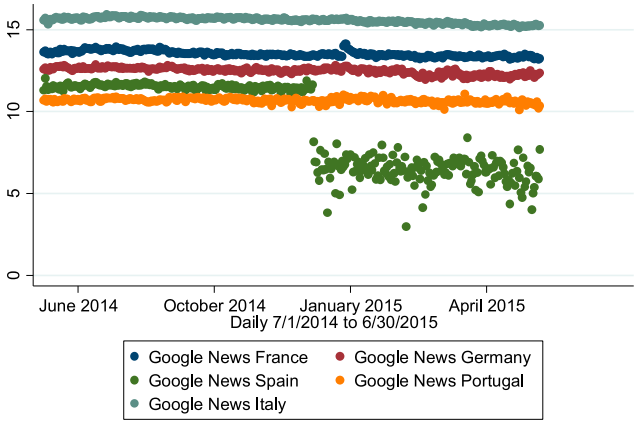
where θ_k is a parameter that captures the net effect of the shutdown on Spanish newspapers in a day t within the kth week after the shutdown of Google News. All other parameters, variables, and fixed effects remain the same from the explanation above. If anything, in our specifications using French 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 that French news outlets received because of 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 that 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 web pages from Spain, France, Italy, Germany, and Portugal.

Figure 1. (Color online) Log Daily Visits to Google News



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

We now describe the overall effect of the Google News shutdown. Table 6 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)–(4) in Table 6 use French outlets as the control group, and columns (5)–(8) in Table 6 use Italian outlets as the control group. The overall findings do not qualitatively change across exercises using different control groups. Columns (1) and (5) in Table 6 show that the shutdown decreased daily visits between 8.4% and 14.6%. See Figure 3 for a graphical representation of the negative effect of the shutdown on the number 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 because of the substitution effect.

In columns (2) and (6) in Table 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 they are consistent with our predictions in Section 3. Before the shutdown, some outlets benefitted from the visits of casual readers generated by Google News (regional and sports outlets), whereas others benefitted less, because they experienced a smaller market expansion effect, which possibly was compensated for by the switch of some (nonloyal) consumers to the aggregator (national and business outlets). These results confirm our prediction that niche outlets and predominant national outlets are less affected by aggregators, because their specialization or ideological bias make them irreplaceable by multihome audiences.

Columns (3), (4), (7), and (8) in Table 6 allow the effect to vary by ranking (top and bottom 50% within our sample according to number of daily visits) and the share of international visitors, respectively. Consistent with the previous results, we find that the impact of the shutdown was larger in lower-ranked domains and domains with lower proportions of international visitors.[29] Figure 4 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 larger shares of international visitors are less likely to be affected by the presence of news aggregators.

A potential concern with our analysis in Table 6 is that the Charlie Hebdo attack of January 7, 2015, and the events that occurred in the following days (police shot on January 8 and a hostage situation at a kosher supermarket near the Porte de Vincennes on January 9) may be driving our results.[31] For this reason, we repeat our analysis in Table 7 after dropping all observations from January 7 to February 4 (four weeks).

Our findings in Table 7 show qualitatively consistent results with those reported in Table 6 using

Figure 2. (Color online) Log Daily Visits to News Aggregators in Spain

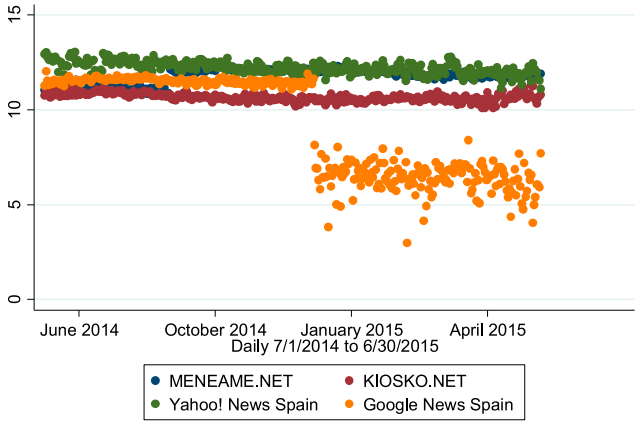


Table 6. Difference in Differences of Shutting Down Google News in Spain after December 16, 2014

	Control group							
	French newspapers				Italian newspapers			
Dependent variable: ln(*Daily visits*)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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^2	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

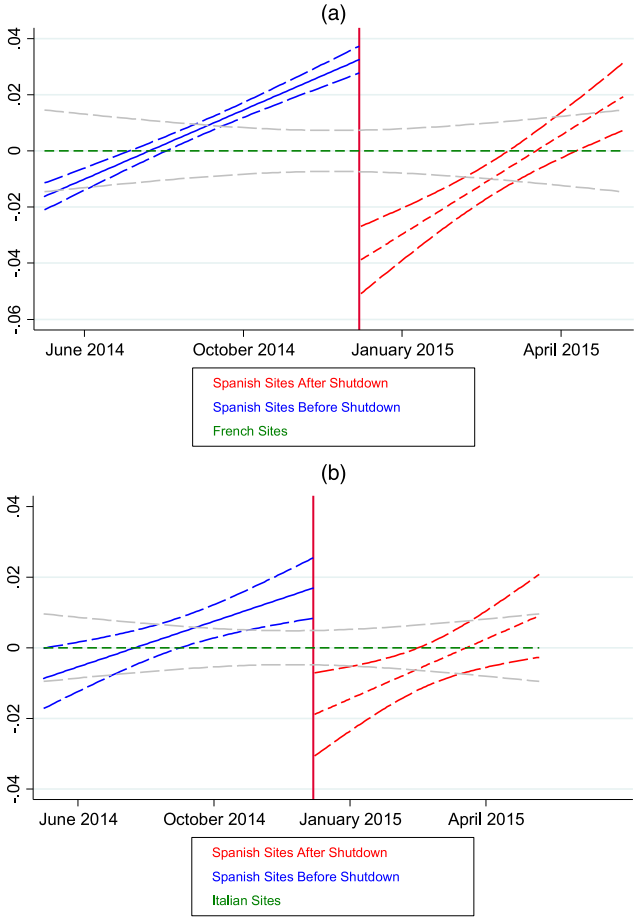
Notes. Robust *t* statistics in parentheses are clustered at the site level. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

French news outlets as the control group (columns (1)–(4) in Table 7). 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 the control group, we obtain qualitatively similar results. Columns (5)–(8) in Table 7 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 Table 7 across different control groups is indicative of the robustness of the result.[33]

Table 8 investigates further the heterogeneity of the effect of the shutdown across outlet types. We run separate analyses for each classification using their respective French and Italian news outlets as control groups in Table 8. Specifically, we compare Spanish national outlets with French national outlets, Spanish regional outlets with French regional outlets, and so on. Columns (1) and (2) in Table 8 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 the control group. Columns (3) and (4) in Table 8 find that there is no effect on business outlets and only an effect on sports when using Italian sites as the 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 owing to differences between the news market in each country and the different impact of the Charlie Hebdo attack in the control group.[34]

When we analyze the impact on top- and bottom-ranked outlets, we observe in columns (5) and (6) in Table 8 that the impact is very similar across types (14% reduction) or larger in bottom-ranked outlets (Table 8(b)).[35] Finally, columns (7) and (8) in Table 8 consider the top and bottom international 50% outlets

Figure 3. (Color online) (a) Log Daily Visits After Shutdown by Spanish Sites Using French Sites as Control Group and (b) Log Daily Visits After Shutdown by Spanish Sites Using Italian Sites as Control Group



in our sample. The effect of the shutdown is larger for outlets with lower shares of internationalization (15.7% versus 11.3% reduction) using French outlets as the control group.[36] 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 fewer casual readers in the domestic and international markets. 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 Table 8(b) with caution, because our definition may explain why we find reductions 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 9 and 10 for the whole sample and each classification separately using French and Italian sites as control groups. Note that all specifications in Table 9 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 9 and 10 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. Although this finding may be surprising at first, this may reflect that, after the shutdown, casual readers remembered for some time some of the outlets that 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 (2014) observe that casual readers directly visit news outlets that they have discovered in the past with the use of aggregators.

Figure 4. (Color online) (a) Log Daily Visits by Top/Bottom-Ranked Spanish Sites Using French Sites as Control Group and (b) Log Daily Visits by Top/Bottom-Ranked Spanish Sites Using Italian Sites as Control Group

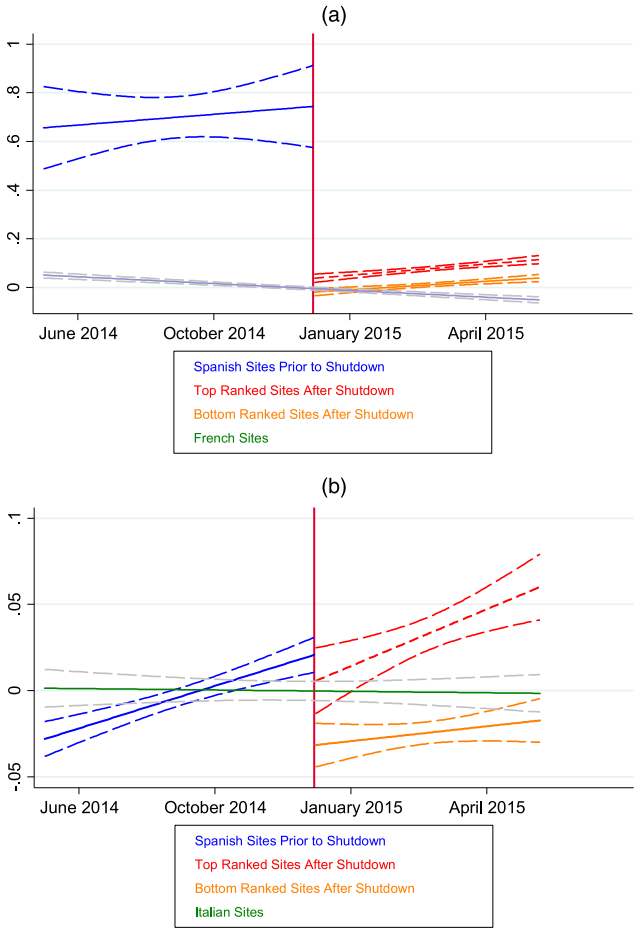


Table 7. Shutting Down Google News Spain and Deleting Four Weeks of Charlie Hebdo Observations

	Control group							
	French newspapers				Italian newspapers			
Dependent variable: ln(*Daily visits*)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Shutdown × *Spain*	−0.0741***				−0.0798***			
	(−2.973)				(−2.699)			
Shutdown × *Spain* × *National*		−0.014				−0.019		
		(−0.345)				(−0.455)		
Shutdown × *Spain* × *Regional*		−0.123***				−0.129***		
		(−3.469)				(−3.315)		
Shutdown × *Spain* × *Business*		0.039				0.033		
		(0.765)				(0.625)		
Shutdown × *Spain* × *Sports*		−0.124***				−0.130***		
		(−3.020)				(−2.949)		
Shutdown × *Spain* × *Top 50%*			−0.006				−0.012	
			(−0.152)				(−0.278)	
Shutdown × *Spain* × *Bottom 50%*			−0.109***				−0.115***	
			(−3.486)				(−3.272)	
Shutdown × *Spain* × *Top int 50%*				−0.023				−0.029
				(−0.611)				(−0.702)
Shutdown × *Spain* × *Bottom Int 50%*				−0.0984***				−0.104***
				(−3.090)				(−2.926)
Constant	14.42***	14.42***	14.42***	14.42***	61.62	63.05	61.59	61.60
	(4.879)	(4.879)	(4.879)	(4.879)	(0.156)	(0.160)	(0.156)	(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^2	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97

Notes. This table repeats the exercise in Table 6 but drops observations from January 7 to February 4, 2015 (both are included). Results are qualitatively the same if dropping from January 7 to January 21 or to January 28, 2015. Robust *t* statistics in parentheses are clustered at the site level. FE, fixed effects.

***$p < 0.01$.

Finally, we test for the "information screening" effect in Tables 11 and 12 by analyzing whether the shutdown of Google News had an impact on the consumers' engagement metrics. To do so, we use seemingly unrelated regressions (suregs) 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 the control group and a decrease in the number of pages per visit when using Italian outlets as the control group (columns (1)–(3) in Tables 11 and 12). We find no differences in changes in engagement metrics between top- and bottom-ranked sites (columns (4)–(6) in Tables 11 and 12). When we decompose the effect by week in columns (7)–(9) in Tables 11 and 12, we find that 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 the Charlie Hebdo attacks occurred when using France as the control group in Table 11.

To interpret this set of results in Tables 11 and 12, 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 at their website via Google News. Second, after the shutdown, some casual consumers could substitute their visits to the news aggregator with those to 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 a 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

Table 8. Difference in Differences by Online Newspaper Type, National Ranking, and International Position

Dependent variable: ln(*Daily visits*)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a): Control group: French newspapers								
Shutdown × *Spain*	−0.210***	−0.179***	0.045	−0.069	−0.144***	−0.141***	−0.113***	−0.157***
	(−6.423)	(−4.375)	(0.748)	(−1.227)	(−3.899)	(−3.439)	(−3.119)	(3–855)
Constant	8.976***	1,439***	−555.500	−3,409**	2,575***	326.5***	−4,704***	21.55***
	(3.163)	(4.277)	(−0.613)	(−2.456)	(4.607)	(3.213)	(−4.096)	(7.461)
Observations	10,585	14,965	2,920	3,650	13,505	18,615	13,505	18,615
R^2	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**	−0.057	−0.052	−0.138***	−0.0636**	−0.102**	−0.100**	−0.076**
	(−2.176)	(−1.423)	(−0.728)	(−4.760)	(−2.138)	(−2.462)	(−2.53)	(−2.26)
Constant	−501.900	−38.510	200 800	−466.700	−159.000	−169.900	−970.397*	−243.530
	(−0.366)	(−0.201)	(0.187)	(−0.779)	(−0.288)	(−0.444)	(−2.002)	(−1.25)
Observations	14,965	16,425	3,650	2,920	15,695	22,265	12,410	25,550
R^2	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%

Notes. All specifications include site and date fixed effects as well as group-specific time trends. Robust *t* statistics in parentheses are clustered at the site level.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

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 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.

4.2.1. 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 who 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 sureg with both daily visits and the consumers' engagement metrics.

Table 13 shows results of our integrated approach using Italian newspapers as the control group.[37] 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) in Table 13 an 8.9% decrease in daily visits associated with the shutdown of Google News in Spain. Columns (2)–(4) in Table 13 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- and bottom-ranked sites, results in columns (5)–(8) in Table 13 show a larger impact on daily visits of bottom-ranked outlets than top-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 (nonsignificant 2.7% versus 2.6%), and bounce rate (0.44% versus 0.46%).

Finally, columns (9)–(12) in Table 13 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 seventh week of the shutdown are consistent with our findings in columns (1)–(4) in Table 13. If anything, we observe gradual increases in visit duration four to six weeks after the shutdown. In the end, the long-term impact on visit duration is null.

Table 9. Difference in Differences of Evolution of Effect by Online Newspaper Type and National Ranking Position

Dependent variable: ln(*Daily visits*)	Control group: French newspapers								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
First week after × Spain	−0.035	−0.0683*	−0.031	0.048	−0.013	−0.030	−0.041	−0.030	−0.018
	(−1.413)	(−1.813)	(−0.740)	(0.796)	(−0.216)	(−0.858)	(−1.281)	(−0.944)	(−0.474)
Second week after × Spain	−0.020	−0.038	−0.055	0.016	0.047	0.025	−0.0759*	0.024	−0.059
	(−0.559)	(−0.791)	(−1.171)	(0.143)	(0.399)	(0.531)	(−1.765)	(0.502)	(−1.393)
Third week after × Spain	−0.0609*	−0.063	−0.087	−0.105	0.041	−0.023	−0.110*	−0.016	−0.091
	(−1.728)	(−1.558)	(−1.453)	(−1.000)	(0.392)	(−0.610)	(−1.868)	(−0.409)	(−1.642)
Fourth week after × Spain	0.220**	−0.079	−0.294***	0.037	0.354***	−37.26***	−2.210**	−13.13***	−0.151*
	(2.345)	(−1.588)	(−3.581)	(0.311)	(5.013)	(−4.581)	(−2.301)	(−4.200)	(−1.686)
Fifth week after × Spain	−0.216***	−0.348***	−0.213***	−0.143**	0.099	−0.257***	−0.148***	−0.208***	−0.181***
	(−4.998)	(−5.393)	(−3.419)	(−3.083)	(1.789)	(−4.495)	(−2.781)	(−3.707)	(−3.102)
Sixth week after × Spain	−0.110***	−0.154***	−0.122**	−0.067	0.021	−0.144***	−0.078	−0.079	−0.101**
	(−3.564)	(−3.608)	(−2.560)	(−1.100)	(0.282)	(−4.450)	(−1.565)	(−1.647)	(−2.297)
Seventh week + after × Spain	−0.137***	−0.182***	−0.153***	−0.119**	0.055	−0.146***	−0.132**	−0.136***	−0.125**
	(−4.386)	(−3.366)	(−3.079)	(−2.320)	(0.728)	(−4.741)	(−2.266)	(−3.327)	(−2.634)
Fourth week after× France	0.770***	0.697***	0.438***	0.181	0.179**	−36.71***	−1.738*	−12.70***	0.493***
	(6.066)	(4.857)	(4.925)	(1.209)	(3.472)	(−4.508)	(−1.815)	(−4.093)	(4.601)
Constant	14.70***	9.024***	1,230***	−3,766**	45.85***	2,575***	311.0**	−4,970***	21.53***
	(4.968)	(3.140)	(3.363)	(−2.570)	(5.681)	(4.610)	(2.420)	(−4.198)	(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^2	0.97	0.97	0.97	0.93	0.97	0.96	0.94	0.97	0.96

Notes. Robust *t* statistics in parentheses are clustered at the site level. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

4.2.2. 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 A.1 in the online appendix for a list of Portuguese news outlets).[38] 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 pretreatment outcome variable. The benefit of building this synthetic control group is that the preshutdown 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).

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 four 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. Although the weights are fixed after the shutdown, the visits to the outlets in the control groups change over time, and therefore, the synthetic control will as well. 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 before treatment, this method creates optimal weights using the number of daily visits (for only a subset of the pretreatment period), share of domestic visits, national rank, and engagement metrics.

Figure 5, (a) and (b) shows the average daily visits for Spanish news outlets and its synthetic counterpart during the period analyzed. Although the exercise in Figure 5(a) considers outlets from all four countries as potential controls (France, Germany, Italy, and Portugal), the analysis in Figure 5(b) 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 Figure 5, (a) and (b), the synthetic control closely tracks the average Spanish

Table 10. Difference in Differences of Evolution of Effect by Online Newspaper Type and National Ranking Position

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

Notes. Robust *t* statistics in parentheses are clustered at the site level. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

news outlets before 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. Similar 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 5(c) shows the results of using all four countries as potential control groups, and Figure 5(d) shows results using national news outlets from only Italy and Portugal. Note that, in both cases, the treatment and the control closely track each other before the shutdown of Google News and that they diverge after the shutdown. These results show that our findings in Tables 6 and 7 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. Although 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 advertisement 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 specializing 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 advertisement inventory. For the purpose of our analysis, we separate these websites into two groups: namely, news outlets and nonnews outlets. Most of the nonnews 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 nonnews sites, and therefore, this control group is an optimal candidate when studying the effect of the shutdown on advertising revenues. Although both news and

Table 11. Seemingly Unrelated Regressions of Difference in Differences Using Engagement Metrics as Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent 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	−0.0376**	0.00424***						
	(0.355)	(−2.325)	(2.765)						
Shutdown × *Spain* × *Top* 50%				0.011	−0.0388**	0.00438***			
				(1.173)	(−2.217)	(2.638)			
Shutdown × *Spain* × *Bottom* 50%				−0.001	−0.0370**	0.00417***			
				(−0.100)	(−2.238)	(2.662)			
First Week After × *Spain*							−0.0746***	−0.125***	0.002
							(−4.497)	(−3.999)	(0.516)
Second Week After × *Spain*							−0.0716***	−0.110***	0.00835***
							(−4.292)	(−3.497)	(2.803)
Third Week After × *Spain*							−0.0701***	−0.198***	0.00926***
							(−4.183)	(−6.275)	(3.098)
Fourth Week After × *Spain*							0.0898***	0.053	−0.00700**
							(4.987)	(1.566)	(−2.178)
Fifth Week After × *Spain*							0.0438***	0.017	−0.002
							(2.587)	(0.519)	(−0.779)
Sixth Week After × *Spain*							−0.020	−0.004	0.002
							(−1.154)	(−0.118)	(0.817)
Seventh Week + After × *Spain*							0.0660***	0.0396**	0.00826***
							(6.343)	(2.027)	(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^2	0.614	0.806	0.761	0.614	0.806	0.761	0.616	0.807	0.761
Breusch–Pagan test	$\chi^2(3) = 10{,}971.572$		Pr = 0.0000	$\chi^2(3) = 10{,}973.458$		Pr = 0.0000	$\chi^2(3) = 10{,}955.74$		Pr = 0.0000

Notes. Robust *t* statistics are parentheses. FE, fixed effects.
**$p < 0.05$; ***$p < 0.01$.

Table 12. Seemingly Unrelated Regressions of Difference in Differences Using Engagement Metrics as Dependent Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent 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***	0.027	0.00449***						
	(–5.323)	(1.593)	(2.968)						
Shutdown × *Spain* × *Top* 50%				–0.0412***	0.026	0.00464***			
				(–4.036)	(1.372)	(2.765)			
Shutdown × *Spain* × *Bottom* 50%				–0.0531***	0.028	0.00442***			
				(–5.594)	(1.586)	(2.842)			
First Week After × *Spain*							–0.0734***	0.003	0.001
							(–4.118)	(0.082)	(0.193)
Second Week After × *Spain*							–0.0961***	0.049	0.0110***
							(–5.364)	(1.499)	(3.743)
Third Week After × *Spain*							–0.0488***	–0.020	0.00592**
							(–2.716)	(–0.616)	(2.008)
Fourth Week After × *Spain*							–0.008	0.0631*	0.004
							(–0.401)	(1.787)	(1.408)
Fifth Week After × *Spain*							–0.021	0.034	0.003
							(–1.155)	(1.031)	(0.932)
Sixth Week After × *Spain*							–0.0404**	0.129***	0.001
							(–2.206)	(3.849)	(0.490)
Seventh Week + After × *Spain*							–0.0439***	–0.0003	0.00450**
							(–3.931)	(–0.0125)	(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^2	0.60	0.78	0.77	0.60	0.78	0.77	0.60	0.78	0.77
Breusch–Pagan test	$\chi^2(3) = 15{,}873.282$		Pr = 0.0000	$\chi^2(3) = 15{,}874.718$		Pr = 0.0000	$\chi^2(3) = 15{,}879.006$		Pr = 0.0000

Notes. Robust *t* statistics are parentheses. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

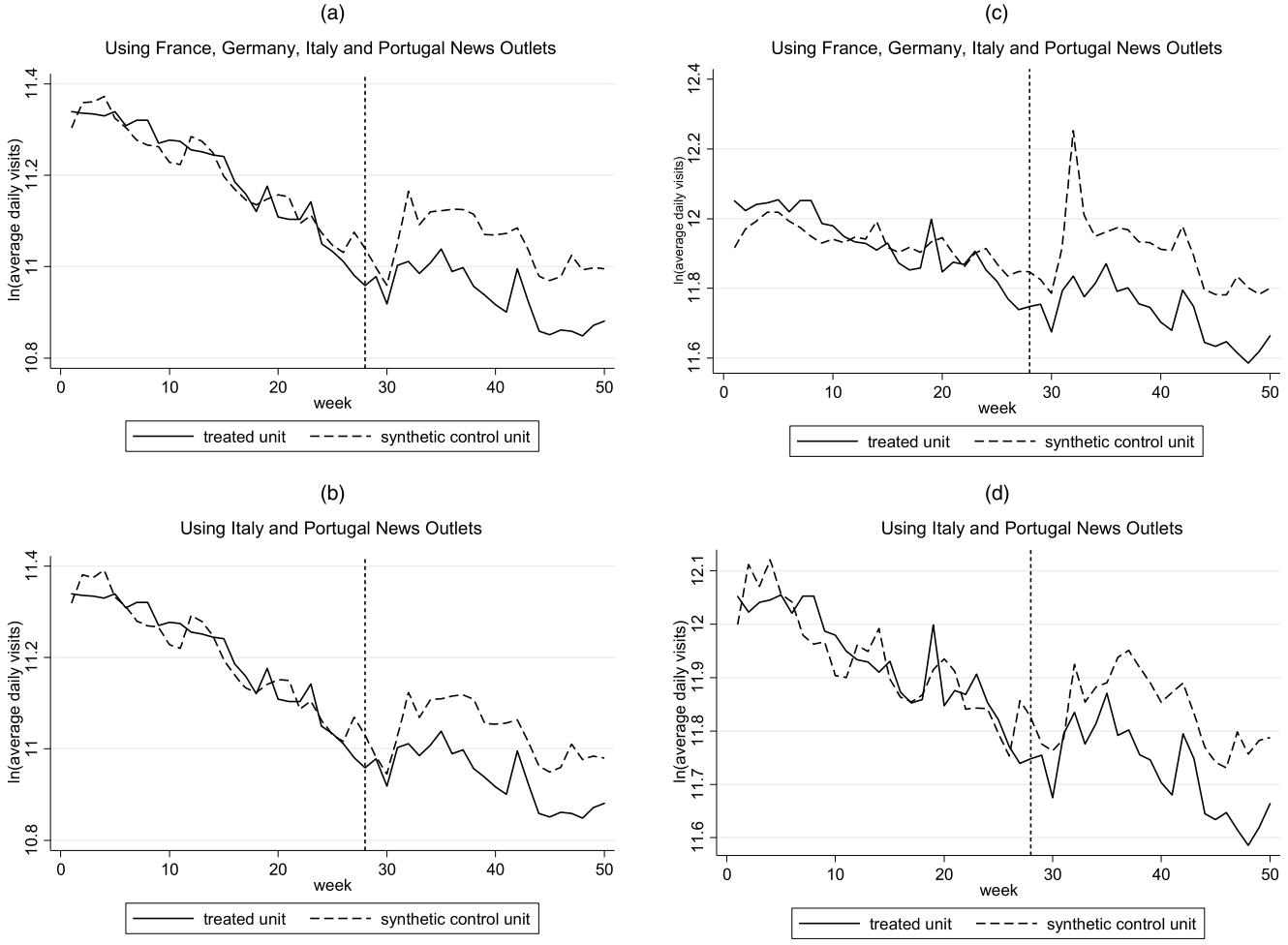
Table 13. Seemingly Unrelated Regressions of Difference in Differences Using Daily Visits and Engagement Metrics as Dependent Variables and Italian News Outlets as the Control Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent 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***	−0.0491***	0.0269	0.0045***								
	(−8.098)	(−5.323)	(1.593)	(2.968)								
Shutdown × *Spain* × *Top* 50%					−0.0329***	−0.0412***	0.0256	0.0046***				
					(−2.690)	(−4.036)	(1.372)	(2.765)				
Shutdown × *Spain* × *Bottom* 50%					−0.118***	−0.0531***	0.0275	0.0044***				
					(−10.38)	(−5.594)	(1.586)	(2.842)				
First Week After × *Spain*									−0.0202	−0.0734***	0.0027	0.0006
									(−0.945)	(−4.118)	(0.082)	(0.193)
Second Week After × *Spain*									−0.0423**	−0.0961***	0.0491	0.0110***
									(−1.971)	(−5.364)	(1.499)	(3.743)
Third Week After × *Spain*									−0.0255	−0.0488***	−0.0202	0.0059**
									(−1.184)	(−2.716)	(−0.616)	(2.008)
Fourth Week After × *Spain*									−0.136***	−0.0078	0.0631*	0.0045
									(−5.871)	(−0.401)	(1.787)	(1.408)
Fifth Week After × *Spain*									−0.112***	−0.021	0.0343	0.0028
									(−5.126)	(−1.155)	(1.031)	(0.932)
Sixth Week After × *Spain*									−0.115***	−0.0404**	0.129***	0.0015
									(−5.224)	(−2.206)	(3.849)	(0.490)
Seventh Week + *After* × *Spain*									−0.137***	−0.0439***	−0.0003	0.0045**
									(−10.27)	(−3.931)	(−0.0125)	(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^2	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	$\chi^2(6) = 16{,}669.243$		Pr = 0.0000		$\chi^2(6) = 16{,}670.633$		Pr = 0.0000		$\chi^2(6) = 16{,}680.946$		Pr = 0.0000	

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

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

Figure 5. (Color online) (a) Synthetic Control Analysis of All News Outlets Using France, Germany, Italy, and Portugal News Outlets; (b) Synthetic Control Analysis of All News Outlets Using Italy and Portugal News Outlets; (c) Synthetic Control Analysis of National News Outlets Only Using France, Germany, Italy, and Portugal News Outlets; and (d) Synthetic Control Analysis of National News Outlets Only Using Italy and Portugal News Outlets



nonnews sites compete for advertisers, only news sites were affected by the shutdown of Google News.

The advertisement data contain information on 78 online domains, 47 news outlets, and 31 nonnews outlets (Table A.2 in the online appendix 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 the website of the news outlet (and the other domains) several times per day, and for each of them, it calculates the number of times that 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 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. After 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" that distributes the remaining advertisement inventory in the market at bulk (see the appendix for a thorough description of the Spanish advertising market). Taking this into

Table 14. Advertising Effects of Google News Shutdown on Online News Outlets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	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	Nonnews-reporting web pages								
Newspaper × Shutdown	−0.106***	−0.028*	−0.078***	−0.136**	0.003	−0.140**	−0.084***	−0.041**	−0.043**
	(−3.620)	(−1.872)	(−3.415)	(−2.039)	(0.342)	(−2.240)	(−3.134)	(−2.106)	(−2.054)
Observations	193,015	193,015	193,015	70,383	70,383	70,383	122,632	122,632	122,632
R^2	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 × Front Page 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

Notes. This table shows a different set of difference-in-differences results using as dependent variables the log of daily advertising revenue per site and advertiser, the advertising intensity per site and advertiser, and the revenues per advertising intensity per site and advertiser on the front page and others. Advertising intensity is a measure of number of advertisements per day, site, and advertiser. We compare online newspapers with other nonnews-related web pages. All specifications control for different time trends across groups. Columns (1)–(3) include *Site × Advertiser × Front Page* and *Date* fixed effects. Columns (4)–(9) use *Site × Advertiser* fixed effects and constrain the sample to observations for the front page or other pages. Robust *t* statistics are in parentheses clustered at the advertiser level. FE, fixed effects.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

account, we expect the shutdown of Google News to reduce the outlets' advertisement revenues because of a decrease in CPM (prices will be lower if the demand for slots and the marginal revenue curve decrease), 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 difference-in-differences methodology used in the paper with two main differences. First, we use Spanish news outlets as the treatment group and Spanish nonnews 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 that occurred in advertising behavior at the site-advertiser-page level within two weeks before and after the shutdown of Google News.

Table 14 shows our first set of results using all Spanish news outlets as the treatment group and nonnews outlets as the control group. Columns (1)–(3) in Table 14 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)–(6) in Table 14 only use front page data and site × advertiser fixed effects, whereas columns (7)–(9) in Table 14 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 difference-in-differences specifications in Table 15. Specifications reported in Table 15(a) use all observations (front page and other pages) with site × page and day fixed effects. We find decreases in the number of daily advertisers, the revenue per advertiser and page, and the revenue per advertising intensity after the shutdown. Table 15, (b) and (c) uses observations from front pages and other pages, respectively, with site and date fixed effects. We find decreases in revenue, advertisement intensity, number of daily advertisers, average revenue per advertiser, and revenue per advertising intensity in front pages and decreases in advertisement intensity and advertisement intensity per advertiser in other pages.

Differences in results between Tables 14 and 15 reveal differences in the effect of the shutdown at the advertiser × site level and the aggregated site level. These differences may reflect a heterogeneous impact of the shutdown across news outlets and

Table 15. Advertising Effects of Google News Shutdown on Online News Outlets

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	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	Nonnews-reporting web pages					
Sample	All observations					
Specification	*Site* × *Front Page* 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^2	0.87	0.97	0.91	0.85	0.93	0.93
			Panel (b)			
Treatment group	Online newspapers					
Control group	Nonnews-reporting web pages					
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^2	0.79	0.88	0.88	0.77	0.66	0.80
			Panel (c)			
Treatment group	Online newspapers					
Control group	Nonnews-reporting web pages					
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^2	0.89	0.89	0.91	0.85	0.84	0.91

Notes. This table shows two different sets of difference-in-differences 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 the number of advertisements per day in a site per advertiser. Panel (a) compares online newspapers with other nonnews-related web pages using all data. Panel (b) uses data only from front pages, and panel (c) uses data only from other pages. Robust t statistics are in parentheses clustered at the site level.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

advertisers.[39] On the one hand, Table 14 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 nonfront pages (4%) and not at all in front pages. These results make sense taking into account that (1) news aggregators link the pages of the news stories and not the front page of the outlets and (2) willingness to pay for advertisements in front pages is larger than in nonfront pages. On the other hand, Table 15(a) 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 14), revenue per advertiser, and revenue per intensity unit. Table 15, (b) and (c) also shows that the intensity per advertiser did not go down in front pages, but it did in nonfront pages (consistent with Table 14). Daily intensity went down in both front and other pages, number of advertisers went down in front pages but not in nonpages, and revenue mostly went down in front pages and did not in nonfront 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. Although we do not observe advertising prices, a possible decrease in prices for nonfront page advertising could help to interpret these findings.

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 the appendix, we detail that, although programmatic advertising is sizable nowadays, at the time of the Google News shutdown, only a relatively small part of the advertisement inventory was commercialized through programmatic advertising. Yet, the presence of targeted advertising poses the questions 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 owing 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, although 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, because the aggregator did not shut down. 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.[40]

The second relevant difference between the Spanish case and the German case is that, in the latter case, the treatment period that we examine took place for a finite amount of time from October 23 to November 5, 2014.[41] After this period, VG Media outlets decided to opt back in to Google News. The short opt-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:

$$\ln[y_{ijt}] = \theta Germany_i \times Opt\_Out_t + \mu_{j0} t_{jt} + \mu_{j1} t_{jt}^2 + \gamma_i + \delta_t + u_{ijt},$$

where all dependent variables are defined as in the previous section. $Germany_t$ is a dummy that equals one if online newspaper i is German and zero otherwise. The dummy Opt_Out_t takes the value of one if day t is between October 24 and November 5, 2014 and zero otherwise. However, 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 identically independently and normally distributed (niid) as usual.

We also consider other specifications in which the treated group included the outlets that opted out. Specifically, we run separate regressions for the members of the VG Media association and 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 it was the first to announce its opt-out choice in October 2014. Finally, we also break the Opt_Out_t dummy into $FirstWeekOpt\_Out_t$ and $SecondWeekOpt\_Out_t$, which takes the value of one 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 because of the competition effect.

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 the control group. Column (1) in Table 16 shows that, after the opt-out decision, the number of daily visits to German outlets did not change relative to Italian outlets. Column (2) in Table 16 shows that this finding is robust when splitting German outlets into VG Media sites and non-VG Media sites. If anything, column (2) in Table 16 shows that the number of daily visits to VG Media outlets experienced a statistically nonsignificant decrease of 2.4%. We focus on Axel Springer sites in columns (3) and (4) in Table 16, and our results show that, although 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 16 use German outlets only and consider the effect of the daily visits on the news outlets that opt out; we used as a control group the rest of German outlets in our sample. Column (5) in Table 16 shows again the existence of a negative but nonsignificant effect of the opt-out decision on the visits to the VG Media outlets

Table 16. Difference in Differences of the Opting Out Period in Germany in October and November 2014 Using Italian Newspapers as the Control Group

	ln(*Daily Visits*)							
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Opt Out × Germany	0.008	0.020	0.032	0.033				
	(0.184)	(0.418)	(0.705)	(0.736)				
Opt Out × VGM		−0.024			−0.024			
		(−0.701)			(−0.688)			
Opt Out × Axel Springer			−0.0763**			−0.0763**		−0.081*
			(−2.506)			(−2.459)		(−1.89)
1st Week Opt Out × Axel Springer				−0.0971**			−0.0814**	
				(−2.621)			(−2.074)	
2nd Week Opt Out × Axel Springer				−0.0740*			−0.0899**	
				(−1.902)			(−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^2	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.01

Notes. Robust t statistics are in parentheses and clustered at the site level. Column (8) reports difference-in-differences estimates with propensity score matching using the 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. FE, fixed effects.

*$p < 0.1$; **$p < 0.05$.

Table 17. Difference in Differences Germany Opt-Out Robustness Check

	ln(*Daily Visits*)			
Dependent variable	(1)	(2)	(3)	(4)
Opt Out × *Axel Springer*	−0.0645*		−0.0830***	
	(−1.964)		(−2.934)	
1st Week Opt Out × *Axel Springer*		−0.070		−0.0881**
		(−1.528)		(−2.495)
2nd Week Opt Out × *Axel Springer*		−0.0784**		−0.0966**
		(−2.117)		(−2.388)
Opt Out × *Top* 50%			0.0457	0.046
			(1.329)	(1.335)
Opt Out × *Regional*	0.0466	0.0466		
	(1.432)	(1.432)		
Opt Out × *Sports*	0.043	0.0441		
	(0.657)	(0.680)		
Opt Out × *Business*	−0.0675	−0.0664		
	(−1.354)	(−1.322)		
Constant	11.08***	11.08***	11.08***	11.08***
	(176.200)	(176.200)	(174.800)	(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^2	0.95	0.95	0.95	0.95

Notes. Robust *t* statistics are in parentheses and clustered at the site level. FE, fixed effects.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

relative to all other German outlets in our sample. In columns (6) and (7) in Table 16, 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 8.1% and 8.9% decreases in the first and second weeks of the opt-out period, respectively. To summarize, our results suggest that 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) in Table 16 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)–(7) in Table 16 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.[42]

A reasonable concern when interpreting our findings in columns (6) and (7) in Table 16 is the endogeneity surrounding the decision of opting out by VG Media and Axel Springer outlets. To address this issue, column (8) in Table 16 shows the results of running propensity score matching between all Axel Springer outlets and the rest of German outlets in our sample as the control group and running difference-in-differences regressions using the closest match as the control group for each treated Axel Springer outlet. Column (8) in Table 16 shows a decrease of 8.1% in daily visits, consistent with other findings in columns (6) and (7) in Table 16.[43] We find qualitatively similar results in Table A.6 in the online appendix when using French news outlets as the control group.[44]

Next, Table 17 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) in Table 16 while also including as independent variables the interaction between the *Opt_Out* dummy variable and the specialization categories of outlets. Columns (1) and (2) in Table 17 take into account the classification of the news outlets according to their content (national, regional, business, and sports), and columns (3) and (4) in Table 17 take into account the classification of the news outlets according to their ranking (top 50%). Results in columns (1) and (2) in Table 17 show that 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) in Table 17 offer similar insights to those in Table 16.[45]

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 nonsignificant effect on their daily visits. Moreover, the effect of this decision was heterogeneous across VG Media' outlets, because 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 longer than average excerpts and the presence of accompanying images increase the probability of links 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."

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 was no competition effect, because 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 multihome and use aggregators to gain access to additional content than that 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 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 that 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 that 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 most traditional publishers unsatisfied; in the last decade, they have seen how the significant increases 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.[46]

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 that 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 nonloyal readers who are aware of the outlets but 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 show how the news outlets consumers' base can determine the benefits that 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 a mechanism to screen news stories and 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 nonfront pages.

Our analysis of the German case has shown that changes in the size of the excerpts or the images that the aggregators release modify the traffic that 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 that these outlets could have received from Google was diverted to other outlets that opted in, which reveals the relevance of news aggregators for competition.

6.1. Limitations

Despite the robustness of our main findings, our study is not free of limitations. On the one hand, our domain-level data do 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. Although 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 do not contain information on advertising prices. Although we restrict our advertising analysis to 15 days before and after the shutdown, we cannot truly disentangle whether differences in advertising revenues per advertiser in a site are caused by changes in prices or the number of daily visits.

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

6.2. 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 through both an increase in traffic and increases in other performance indicators, such as advertising revenues, advertising slots, and the number of advertisers. Our findings have important implications for policy makers interested in understanding the impact of copyright legislation that may limit the aggregators' access to the publishers' content.

Although our research answers the question of whether news aggregators are predominantly increasing the audience and market reach of news outlets, we believe that 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.[47] A full understanding of the effect of news aggregators on news consumption is essential for copyright policy design that benefits consumers and societies overall.

Acknowledgments

The authors thank Shane Greenstein, Avi Goldfarb, Hal Varian, Susan Athey, Omar Nayeem, Mike Bailey, Lisa George, Paul Seabright, Lesley Chiou, Gerard Llobet, Jorge López, and Alvar Teixidor for their valuable comments and suggestions. The authors also thank audiences at the Winter

National Bureau of Economic Research Meeting on the Economics of Digitization; the Eighth Annual Conference on Internet Commerce and Innovation at the Searle Center of Northwestern University; the New Economic School in Moscow; the Johns Hopkins Carey Business School; the American University; the Federal Communications Commission; the Universitat de Barcelona; the Universitat Oberta de Catalunya; the 10th Conference on "The Economics of Intellectual Property, Software and the Internet" at Toulouse; the Florence School of Regulation Communications & Media; the Centre for Media Pluralism and Media Freedom and European Networking and Training for National Competition Enforcers Annual Scientic 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.

Appendix. Internet Advertising Market in Spain

The internet advertising market in Spain underwent an important transformation during the time period of our study not only because of the use of new information management technologies but also, because of the appearance of several new agents that facilitated the coordination of publishers and advertisers. Indeed, in recent years, direct contracting to sell advertisement 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. Advertising networks buy unsold inventory from publishers that they later resell through the use of technologies that help with categorizing, packing, and selling such slot inventory. Advertising exchanges facilitate the commercialization of advertisement inventory from multiple advertising networks. Instead of negotiating prices, buyers and sellers of slots may predetermine prices and audience characteristics that they are interested in (many consider advertising exchanges as the origin of programmatic advertisement). Additionally, both publishers and advertisers may use advertising 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 advertisement inventories. Programmatic advertising is now able to efficiently manage buying and selling advertisements 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 is an auction-based mechanism by which advertisement inventory is bought and sold on a per impression basis via programmatic instantaneous auction. Buyers bid on an impression, and if they win, their advertisement is instantly displayed on the publisher's site. Advertising exchanges are auction-based marketplaces that facilitate this process. (2) Programmatic direct is a process that permits advertisers to reserve an advertisement slot that they will use in the future while fixing in advance the price and the characteristics of the targeted audience of their advertisement campaign. Although 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 advertisement slot, news outlets prioritize their use. The remaining unsold advertisement inventory is then offered through programmatic advertisement. Although news outlets benefit from programmatic advertisement in that they are now able to monetize all of their inventory, they have little market power, and the prices that 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 levels 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 advertisement in the web page, exposure time, or the impact of the advertisement campaign. The prices also depend on the type of advertisement, which differs between standardized banners, popups, and videos. There are also several ways to contract for advertisements: (1) CPM, (2) cost per click, (3) cost per contact, person (used to create a database of potential consumers), or (4) cost per action, which can be one click, a subscription, or a sale. 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, it represented 37% in France, and it represented 30% in Italy.[48] According to IAB (2014), programmatic advertisement represented about 20% of the total spending on digital publicity in Spain.[49]

Endnotes

[1] Google launched Google News in 2002, and today, it 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.

[2] See Goldfarb and Tucker (2019) for a review of the literature analyzing the effects of copyright laws in digital markets.

[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 et al. 2017a).

[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 the news release at 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 from use of its contents, arguing that Google was refusing to pay for the links and driving traffic away from their websites. In the following years, some news outlets allowed back links from the aggregator. See the news release at 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 antitrust authorities in 2011. See the news release at http://www.nytimes.com/2011/01/18/technology/18iht-google18.html?_r=2.

[7] See the news release at http://www.theverge.com/2012/12/13/3764692/google-copyright-lawsuit-settlement-belgium.

[8] See the news release at http://googlepolicyeurope.blogspot.de/2012/12/partnering-with-belgian-news-publishers.html.

[9] See the news release at https://googleblog.blogspot.fr/2013/02/google-creates-60m-digital-publishing.html.

[10] See the news release at http://googleespana.blogspot.com.es/2015/04/google-y-editores-de-medios-de-europa.html.

[11] See the news release at 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 Asociación Española de Editoriales de Publicaciones Periódicas and Coalición Pro-Internet, opposed it. The Spanish regulator Comision Nacional de los Mercados y la Competencia also advocated for the modification of several aspects of the new regulation. See CNMC (2014) and Llobet (2015).

[13] See the news release at http://googleespana.blogspot.com.es/2014/12/novedades-acerca-de-google-noticias-en.html.

[14] See the news release at https://support.google.com/news/answer/6140047?hl=es.

[15] See the news release at http://www.aede.es/wp-content/uploads/2015/02/AEDEPrensa-CierreGoogleNewsDic14.pdf.

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

[17] VG Media accused Google of "extortion" and "discrimination" against its members, saying that the law passed the previous year guaranteed news providers a "reasonable compensation for the use of their content." See the news release at 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 the news release at 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 at http://www.axelspringer.de/en/presse/Axel-Springer-concludes-its-data-documentation-Major-losses-resulting-from-downgraded-search-notices-on-Google_22070687.html.

[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 the news release at http://www.ibtimes.com/news-media-no-match-google-axel-springers-about-face-1720494.

[22] Regional outlets have high loyalty and high brand awareness among regional readers, but on the national stage, there is likely less brand awareness; therefore, 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 the team in that region, brand awareness may be high, but at a national level, it is low. Consumers who multihome only browse one sports outlet (even if they know about the existence of others), but when they have access to 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 web page:

> 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 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 information on Joel Sommerland in *The Independent* on June 18, 2018, at 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 fewer visits than news outlets in Spain, France, and Italy because of the smaller population size in Portugal. SimilarWeb data are less precise when the number of daily visits is below 5,000, and therefore, the number of Portuguese news outlets with reliable daily information is far smaller than that in other countries, such as Italy and France.

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

[26] Despite the shutdown of Google News in Spain, the domain kept having a small number of visitors, because its web page explains Google's shutdown decision. See Figure A.1 in the online appendix for the default web page of news.google.es still posted.

[27] Figure 3 is the result of a multistage process. First, we run an ordinary least squares (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. Fourth, we fit lines for Spain before and after the shutdown and another line for the control group [French outlets for Figure 3(a) and Italian outlets for Figure 3(b)].

[28] When testing statistical differences across coefficients in column (2) in Table 6, 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. We reject equality across all pairwise comparison of coefficient except for the pair of national and business coefficients, and the pair of regional and sports coefficients. We find the same patterns of statistical differences across parameters in column (6) in Table 6 when using Italian outlets as control group.

[29] When testing statistical differences across coefficients in columns (3) and (4) in Table 6, we are able to reject equality of coefficients of top- and bottom-ranked outlets at 2%, and we are also able to reject equality of coefficients of top- and bottom-ranked international

outlets at 7%. We find the same results when comparing coefficients across groups for columns (7) and (8) in Table 6.

[30] Similar to Figure 3, Figure 4 is the result of a multistage 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. Fourth, we fit lines for Spain before the shutdown and top- and bottom-ranked Spanish sites after the shutdown and another line for the control group (French outlets for Figure 4(a) and Italian outlets for Figure 4(b)).

[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 difference-in-differences 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 able to control for.

[32] In columns (2) and (6) of Table 7, we cannot reject equality of coefficients of national and business outlets, and regional and sports outlets, respectively. Additionally, we cannot reject equality of coefficients of top and bottom half international outlets in columns (4) and (8) (p-value at 12%) of Table 7. We can reject the null hypotheses of equality for all other pairwise combinations of coefficients in Table 7.

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

[34] We find no statistically significant difference when testing the national and regional coefficients across regressions in Table 8(a). 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 is not conservative enough when finding statistically significant differences if we expect standard errors to increase if clustering at the site level. When using Italian news outlets as control groups (Table 8(b)), we are only able to reject equality between coefficients in columns (2) and (3) in Table 8 at the 9% level. All other coefficient pairs show no statistically significant difference.

[35] We find no significant statistical difference between coefficients across regressions in column (5) and (6) in Table 8 when using French or Italian outlets as the control group.

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

[37] Table A.3 in the online appendix shows qualitatively similar results when using French news outlets as the control group.

[38] Because Portuguese news outlets have a much smaller number of daily visits than Spanish, French, and German outlets, the data from SimilarWeb on Portuguese outlets are noisier and less precise than those for outlets from other countries.

[39] Tables A.4 and A.5 in the online appendix shows results of running sureg. Although 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 the site-page type-day level. These specifications allow for the error term to be correlated across regressions.

[40] Note that the German case does not allow us to identify separately the competitive effect from the "information screening effect" introduced in Section 3. Briefly, the adoption of the opt-in policy did not prevent Google from indexing news from opt-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.

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

[42] 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 10 outlets of 16. The six VG Media's outlets not part of Axel Springer are regional outlets with a low domestic ranking and business and sports outlets.

[43] 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 percentage domestic visits. We perform the difference in differences only on the common support of the propensity score with the kernel option.

[44] If anything, columns (1)–(4) in Table A.6 of the online appendix finds that German outlets, on average, received higher numbers of daily visits than French news outlets. In our opinion, this result may be owing to the sharp increase and posterior decrease in daily visits because of 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.

[45] Table A.7 in the online appendix examines the effects of Axel Springer's opt-out decision on the engagement metrics. We do not observe statistically significant changes.

[46] See https://elpais.com/elpais/2017/03/24/inenglish/1490362000_295766.html.

[47] In a recent paper, Angelucci and Cagé (2019) 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 led to a decrease in advertising revenues of national newspapers compared with 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 the entry of television broadcast national news in the 1950s in the U.S. news market.

[48] See Grece (2016).

[49] See IAB (2014).

References

Abadie A, Gardeazabal J (2003) The economic costs of conflict: A case study of the Basque country. *Amer. Econom. Rev.* 93(1):113–132.

Abadie A, Diamond A, Hainmueller J (2015) Comparative politics and the synthetic control method. *Amer. J. Political Sci.* 59(2):495–510.

Angelucci C, Cagé J, Sinkinson M (2018) Media competition and news diets. Accessed April 23, 2019, https://sites.google.com/site/charlesangelucci/research.

Angelucci C, Cagé J (2019) Newspapers in times of low advertising revenue. *Amer. Econom. J.: Microeconom.* Forthcoming.

Athey S, Mobius M (2012) The impact of news aggregators on internet news consumption: The case of localization. Working paper, Stanford University, Stanford, CA.

Athey S, Calvano E, Gans J (2017a) The impact of consumer multihoming on advertising markets and media competition. *Management Sci.* 64(4):1574–1590.

Athey S, Mobius M, Pal J (2017b) The impact of aggregators on internet news consumption. Working paper, Stanford University, Stanford, CA.

Cagé J, Herve N, Viaud M-L (2015) The production of information in an online world. Working Paper 15-05, NET Institute, New York University, New York.

Calzada J, Ordóñez G (2012) Competition in the news industry: Fighting aggregators with versions and links. Working Paper 12-22, NET Institute, New York University, New York.

Chesnes M, Dai W, Jin GZ (2017) Banning foreign pharmacies from sponsored search: The online consumer response. *Marketing Sci.* 36(6):879–907.

Chiou L, Tucker C (2017) Content aggregation by platforms: The case of the news media. *J. Econom. Management Strategy* 26(4): 782–805.

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. Working Paper PRO/CNMC/0002/14, Comisión Nacional de los Mercados y la Competencia, Madrid.

Dellarocas C, Katona Z, Rand WM (2013) Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Sci.* 59(10):2360–2379.

Dellarocas C, Sutanto J, Calin M, Palme E (2016) Attention allocation in information-rich environment: The case of aggregator. *Management Sci.* 62(9):2543–2562.

George L, Hogendorn C (2012) Aggregators, search and the economics of new media institutions. *Inform. Econom. Policy* 24(1): 40–51.

George L, Hogendorn C (2014) Local news online: Aggregators, geotargeting and the market for local news. Hunter College Department of Economics Working Paper 443, Hunter College, New York.

Goldfarb A, Tucker C (2019) Digital economics. *J. Econom. Literature* 57(1):3–43.

Grece C (2016) The online advertising market in the EU—Update 2015 and focus on programmatic advertising. Report, European Audiovisual Observatory, Strasbourg, Germany.

Huang J (2017) Should Google charge like a taxi driver? Working paper, CEMFI, Madrid.

IAB (2014) El Libro Blanco de la Compra Programática. Accessed April 23, 2019, https://iabspain.es/investigacion/busca-tu-estudio/.

Jeon DS (2018) Economics of news aggregators. Ganuza JJ, Llobet G, eds. *Economic Analysis of the Digital Revolution, FUNCAS Social and Economic Studies, 5* (FUNCAS, Madrid), 343–366.

Jeon DS, Nasr N (2016) News aggregators and competition among newspapers in the internet. *Amer. Econom. J. Microeconom.* 8(4): 91–114.

Llobet G (2015) Los agregadores de noticias y la prensa escrita en la era de Internet. *Papeles de Economía Española* 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). Accessed April 23, 2019, http://www.aeepp.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). Accessed April 23, 2019, https://www.aeepp.com/pdf/Informe_NERA_para_AEEPP_(INGLES).pdf.

Roos J, Mela CF, Shachar R (2018) The effect of links and excerpts on internet news consumption. Accessed April 23, 2019, http://www.jasonmtroos.com/publication/links-excerpts-news/.

Sismeiro C, Mahmood A (2018) Competitive versus complementary effects in online social networks and news consumption: A natural experiment. *Management Sci.* 64(11):5014–5037.