The Impact of the GDPR on Content Providers

V. Lefrere,[*]L. Warberg,[†]C. Cheyre,[‡]V. Marotta,[§]and A. Acquisti[¶]

December 2021

Abstract

While the European General Data Protection Regulation (GDPR) has received considerable attention in economic research, claims and concerns that it would adversely affect websites' ability to provide high quality content to visitors have not been vetted or investigated. In a longitudinal study spanning data collected before and after the GDPR enactment, we track how online content providers (such as News and Media websites) adapted their response to the regulation over time, and whether restrictions on online tracking enforced by the regulation ultimately affected websites' downstream variables such as the quantity of content they offer to their visitors and visitors' engagement with such content. We find evidence of websites' reactions to the GDPR in both the US and the EU, including an initial reduction in the number of third-party cookies and amount of visitors tracking. However, those initial reductions are followed, several months after the enactment of the regulation, by a reversal of the trend and an uptick in tracking among EU websites. We use difference-in-differences, LATE, and look ahead matching specifications to assess downstream effects, distinguishing between ecosystem effects (which affect all EU-based websites relative to US-based websites) and website-level effects (which depend on individual websites' specific responses to the GDPR). We document a reduction in average views per visitors in EU websites relative to US websites, but no statistically significant impact of the regulation on EU websites' new content, ranking, and social media engagement with new content.

[*]Institut Mines Telecom, Business School. Email: vincent.lefrere@imt-bs.eu
[†]Engineering and Public Policy, Carnegie Mellon University. Email: warberg@cmu.edu
[‡]Department of Information Science, Cornell University. Email: ccheyre@infosci.cornell.edu
[§]University of Minnesota Twin Cities Carlson School of Management. Email:vmarotta@umn.edu
[¶]Heinz College, Carnegie Mellon University. Email:acquisti@cmu.edu.

1 Introduction

In May 2018, the European Union (EU) implemented the General Data Protection Regulation (GDPR) to increase the protection of individuals' privacy and their control over personal data. The enactment sparked interest among academics, policy-makers, and industry actors worldwide. Since then, much empirical attention has been devoted to measuring websites' compliance with the GDPR, documenting changes in online consent mechanisms, and estimating compliance costs (Section 2). Less attention has been devoted to understanding downstream consequences that the regulation might have on economically important metrics, such as the ability of websites to produce content, and the ability of Internet users to access and benefit from it. In a longitudinal study spanning data collected before and after the enactment of the GDPR, we track how online content providers (News and Media websites) adapted their response to the GDPR over time, and whether restrictions on online tracking enforced by the regulation ultimately affected websites' downstream variables, such as the quantity of content they offer to their visitors and visitors' engagement with such content.

A defining characteristic of the GDPR is the restrictions it places on the collection and processing of European (EU) residents' data by organizations—including the requirement that organizations justify data collection and processing under various allowed lawful bases. In addition, the GDPR establishes steep financial penalties for organizations that do not comply. For example, the GDPR enabled the Luxembourg National Commission for Data Protection (CNPD) to impose on July 16, 2021 a record fine of €746 million on Amazon for processing personal data, in relation to online tracking and advertising, in a way that did not comply with GDPR (a decision that Amazon called without merit and vowed to vigorously defend against)[1], and the French privacy regulatory authority (CNIL) to fine Google €50 million on January 21,

[1] `https://www.sec.gov/ix?doc=/Archives/edgar/data/0001018724/000101872421000020/amzn-20210630.htm` pp.13

2019 for "lack of transparency, unsatisfactory information and lack of valid consent for the personalization of advertising."[2]

The requirements introduced on the collecting and processing of personal data, and its associated penalties for non-compliance, can affect industries and markets which rely heavily on personal data and profiling. Given how widespread the collection of and reliance on personal data is across different sectors, the overall economic impact of the GDPR was estimated, in early industry reports, to be substantial and negative. For instance, a 2013 Deloitte impact assessment report suggested that the potential economic impact of the GDPR (Deloitte, 2013) could amount to a loss of around 2.8 million jobs and a reduction of European GDP by around 1.34% (corresponding to around €173 billion). The online advertising industry was expected to be especially affected by the GDPR, since its growth is driven by the ability to track users' online behavior to deliver personalized advertising.[3] First, the imposition of limitations on the ability to collect and use data could have a negative impact on the effectiveness of online advertising campaigns (Goldfarb and Tucker, 2011). Second, the GDPR may also impact the composition of the online advertising industry: regulation could impose comparably greater constraints and risks on small and medium-sized online advertising firms, leading to a further concentration of an industry where dominant players already have substantive power to define how the market operates and how benefits are allocated (Johnson and Shriver, 2019). This could in turn expose users, as well as firms upstream and downstream from advertising platforms, to other types of harms, such as monopolistic behavior.

Reductions in advertising effectiveness, spending, and competition were expected,

[2] `https://www.cnil.fr/fr/la-formation-restreinte-de-la-cnil-prononce-une-sanction-de-50-millions-deuros-lencontre-de-la`

[3] For instance, in a recent study (which did not focus on GDPR, but on the impact of reductions in consumer tracking), Johnson *et al.* (2020) found that reductions in the ability to target advertising (through industry self-regulatory initiatives allows American consumers to opt-out from tracking) resulted in a decrease of around $8.58 of ad spending for each consumer who chose to opt out, borne by publishers and ad exchanges.

in turn, to ultimately negatively affect publishers. Since advertising is a major revenue component for digital goods producers (Lambrecht *et al.*, 2014), constraints on online tracking and consumer data gathering may threaten the subsistence of free online content and services. In fact, prior to the enactment of the GDPR, claims by both industry groups and think tanks asserted this hypothesis. In a late 2015 report by IHS Technology, the CEO of the Interactive Advertising Bureau of Europe, Townsend Feehan, suggested that overly burdensome privacy regulation may "limit digital advertising's ability to continue to deliver a wide range of online content to users at little or no cost at the point of consumption" (IHS Technology, 2015). An earlier report by the Information Technology and Innovation Foundation made stronger claims, stating, "[t]he evidence clearly suggests that the tradeoffs of stronger privacy laws result in less free and low-cost content and more spam (i.e. unwanted ads) which is not in the interests of consumers" (Castro, 2010). Both sources capture the sentiment that overbearing privacy regulation could negatively impact publishers, resulting in a reduction in the availability of content or a degradation of its quality.

Despite the numerous claims and predictions about the potential effects of the GDPR on the profitability of ad-supported content providers and services, empirical evidence is limited and contradictory. For instance, anecdotal evidence suggests that at least some online publishers that reduced their use of behaviorally targeted advertising in the EU, post-GDPR, have continued to enjoy stable advertising revenue (Davies, 2019). Additionally, it is unclear how consistently different companies interpreted and applied the regulation. Small firms could find implementing the GDPR costly, and their size may not justify the investment necessary to use personal data in a compliant way. Larger technology firms such as Microsoft and Apple may exploit data protection to achieve competitive advantage (both firms declared before the enforcement of the GDPR that they would voluntarily implement GDPR protections worldwide (Brill, 2018; Phelan, 2018)). Thus, the ultimate implications of the GDPR on the ad-supported

publishing ecosystem may be more nuanced than the negative scenarios being proposed by the online advertising industry. Furthermore, no empirical study has yet directly tested the relationship between the regulation and websites' ability to provide high quality content to visitors have. Understanding the impact of the GDPR on online content providers (including, but not limited to, news websites) and their visitors, and specifically websites' ability to provide content (including free content), is the focus of this manuscript.

We collect data for 909 content providers—news and media websites in both the European Union (EU) and the United States (US). Before and after the enactment of the GDPR, we regularly mine website-level data, from both EU and US IP addresses. The data collected spans a period of time of at least 19 months for some metrics (from April 2018 to November 2019), and much longer for other metrics (April 2017 to November 2019). This longitudinal panel captures how EU and US websites responded to the GDPR over time, including how they interacted over time with visitors and how they managed the collection of visitors' information. We refer to these variables as technical variables in our analysis (Section 4.2). In addition, we use multiple available sources to capture how websites changed their content offerings over time, and how visitors reacted to websites' content over time. We refer to these variables as downstream economic variables in our analysis (Section 4.2).

We focus on content providers as they consist, predominantly, of websites that rely heavily on online advertising. We consider and compare both EU and US websites as the GDPR applies to every European organization (regardless of country of origin of the organization's customers/users), but provides protection to every European resident regardless of whether the organization providing the service is based in Europe or not. That means that websites located in the US should, in principle, comply with the GDPR when interacting with EU visitors. However, the impact of the regulation might be expected to be more evident for European websites, whereas non-European

websites (including US websites), whose user bases are largely non-European, may be only marginally affected.

In our empirical analyasis we distinguish between ecosystem effects (affecting all EU-based websites relative to US-based websites) and website-level effects (which depend on individual websites' specific responses to the GDPR). We first use a difference-in-differences approach to estimate how the regulation impacts EU websites, compared to US websites. To further address endogeneity concerns, we complement that analysis with local average treatment effect (LATE) specifications. Finally, in order to be able to isolate website level effects we used a look ahead matching analysis. This analysis is based on comparing the outcomes experienced by websites that adopt the same response to GDPR, but at different points in time, allowing us to exploit the temporal variation in adoption the identify the effect of the response on outcomes.

The longitudinal nature of our dataset, and the fact that our data include both websites' responses and downstream economic effects, allow us to paint a rich picture of the evolution of the online publishing ecosystem post-GDPR, and to consider the economic impact of the GDPR in light of websites' different choices of response to the regulation. We find robust evidence of websites' reactions to the GDPR in both the US and the EU. We find evidence of significant heterogeneity of responses both between EU and US websites, and also within EU websites. We also find evidence of changes—especially among EU websites—in responses over time. In particular, we find evidence of reduction in the number of third-party cookies and visitors tracking among both EU and US websites. However, those initial reductions are followed, several months after the enactment of the regulation, by a reversal of the trend and an uptick in tracking among EU websites. In particular, we find that websites with the stronger responses to the GDPR are websites that receive only a small fraction of traffic from EU visitors, and thus do not rely on the latter for economic success. Websites that, instead, receive a significant proportion of traffic from EU visitors either do not implement measurable

changes (for instance, by invoking legitimate business interest in the collection of visitors' data), or adjust their responses over time. Accordingly, our various econometric specifications are consistent in failing to reject the null hypothesis of no significant differences in downstream economic outcomes for EU and US websites: while we find a reduction in average views per visitors in EU websites relative to US websites, we find no statistically significant impact of the regulation on EU websites' ability to provide content, on the amount of visitors' traffic they receive, and on visitors' social media engagement with new content. In short, the results seem to suggest that websites that did respond more strongly to the GDPR were the ones that were not likely to be affected by such response; whereas websites that did rely in great part on EU visitors found, over time, ways to avoid being negatively affected by the regulation.

2 Literature Review

This paper builds upon and contributes to three strands of literature: the literature on the economics of privacy (and, in particular, the economic impact of privacy regulation); the growing body of economic and non-economic work on the impact of the GDPR; and the economic literature on the online advertising industry and the media industry, including content providers and online publishers.

Privacy regulation and economic outcomes. The economics of privacy literature investigates the trade-offs associated with the revelation or protection of personal information (Acquisti *et al.*, 2016). Within this literature, an important strand of work has focused on the impact of privacy regulation. Policy interventions that regulate the collection or usage of consumer data tend to be aimed at protecting individuals' privacy, but may have nuanced and unpredictable consequences for innovation, market structure, and the economic welfare of different stakeholders. For instance, Goldfarb and Tucker (2012) argue that privacy regulation might affect the extent and direction

of data-based innovation. Empirical works have showed that the impact of privacy regulation can be quite heterogeneous and context specific. In the health care domain, Miller and Tucker (2009) have found that privacy legislation primarily can reduce demand for electronic medical records (EMR) via a suppression of network effects. On the other hand, Adjerid *et al.* (2015) find that, although privacy regulation can result in a reduction in health information exchanges (HIE)'s operational effectiveness, if the right privacy incentives are provided to patients, regulation can also have a *positive* impact on the development and adoption of HIEs.

The online advertising market is an obvious candidate for the study of how limits imposed on the type or amount of data that can be collected and used may affect an industry reliant on these data. In online advertising, ads are increasingly often targeted to individuals based on information tracked and collected online. Personalized (targeted) ads are likely to be more effective than non-targeted ads since consumers see only those ads that are relevant to them, and this provides higher returns for the advertising companies (Evans, 2009). In fact, Goldfarb and Tucker (2010) empirically investigated how the 2002 EU Privacy and Electronic Communications Directive, which restricted advertisers' ability to collect data on users, affected advertising effectiveness captured by hypothetical purchasing intentions. Their results show that after the regulation, certain types of display advertising were less effective relative to display advertising in other countries. Accordingly, as noted in the Introduction, the advertising industry has been quick to complain that restrictions on their ability to collect and use consumer data for targeted advertising may be harmful to both advertising companies and consumers, as they would impair websites' ability to provide quality content to their visitors (Castro, 2010; IHS Technology, 2015). To our knowledge, however, the link between privacy regulation more broadly (and the GDPR specifically) and websites' ability to provide content has not been vetted in empirical research.

Economic studies of the GDPR. Within the large body of work on privacy

regulation, our paper builds upon the recent wave of empirical studies of the impact of the GDPR. One of the very first studies in this stream by Jia *et al.* (2021) investigated the impact of the GDPR on investments in EU emerging technologies. The authors found that the regulation led to a decrease in such investments for EU companies, compared to US organizations. Similarly, Goldberg *et al.* (2021) examined the effect of the GDPR on European web traffic and e-commerce sales and found that recorded page-views and recorded revenues both fall. The authors observed a drop in page views of around 0.4% and a drop in revenue of around 0.6% for EU users in 2018 compared to 2017. On the other hand, in a theoretical paper, Lefouili and Toh (2018) argued that the effect of the GDPR on investment may be mixed. For example, they suggested that, in a fully covered market, regulating information might reduce investments and yet may be socially desirable when information and quality are not strong complements. In another theoretical piece, Choi *et al.* (2019) investigated consumers' privacy choices with a model in which consumers are required to consent to the collection of their data and consumers are fully aware of the consequences of giving such consent. The authors found that information externalities and coordination failures among users are drivers of excessive loss of privacy. The possibility of highly nuanced and contextual effects of the GDPR emerging from the theoretical literature is consistent with some empirical studies. Zhuo *et al.* (2021) measure the impact of the GDPR on interconnection agreements between EU network providers with those outside the EU. While the authors note a decrease in demand for data within EU networks, they estimate zero effects of the regulation on the number of types of interconnection agreements in the short-run.

Recent work has examined how the GDPR has shaped the advertising market. In a theoretical paper, Sharma *et al.* (2019) argued that burdens imposed by regulations such as the GDPR would negatively impact the revenues derived from ad networks for smaller publishers more than larger publishers. In an empirical investigation, Peukert *et al.* (2020) found a decrease in connections to third-party domains among $110,000$ websites

following the introduction of the GDPR. The authors observed increased concentration among web technology providers as the market share for small firms decreases, while large firms such as Google see significant increases. Johnson and Shriver (2019) found a similar increase in concentration among web technology vendors from a sample of $27,000$ websites.

The impact of the GDPR on websites' interface features (including consent mechanisms, and visitors' reactions to them) has also been investigated, in particular within economics and computer science. Shorty after the enforcement of the GDPR, Libert *et al.* (2018) measured a 22% drop in third-party cookies on news websites. Later, Dabrowski *et al.* (2019) browsed websites from EU and US visitor addresses and found that EU-based visitors were less likely to receive persistent cookies compared to US visitors, even as the number of US-based visitors decreased. In the same vein, Urban *et al.* (2020) show that a particular type of cookie—syncing cookies that allow the exchange of users' information between online advertising actors such as Ad networks and Ad exchanges—decreased across more than 2.6 million websites by approximately 40% around the time the GDPR came into effect. However, the authors found that the number of syncing cookies slightly increase again over the long-term. In a related longitudinal study, Sørensen and Kosta (2019) assessed the effect of the GDPR on the presence of third parties on EU websites. While they found that the number of third parties did slightly decline after the GDPR, they ultimately concluded that the GDPR may not necessarily be responsible for that effect. Degeling *et al.* (2019) investigated online websites' compliance with the data collection requirements imposed by the GDPR. They found that while most websites adjusted their privacy policies and implemented consent mechanisms in the months immediately following GDPR enforcement, some had not complied and did not provide users with means to meaningfully consent to tracking. In an empirical investigation of intermediaries in the online travel industry, Aridor *et al.* (2020) found that the total number of consumers observed by the inter-

mediary decreased by 12.5% after the GDPR, suggesting that a significant number of consumers decided to opt-out. The authors additionally found that the remaining set of consumers who decided to not opt-out were more persistently identifiable. Finally, they observed a drop in ad interactions across their data-set, along with an increase by advertisers in the average bids for the remaining observable consumers, leading to a smaller overall decline in revenue. In a separate study, Sanchez-Rola *et al.* (2019) investigated the use of opt-out options by users and found that, despite the presence of the opt-out mechanism, it was still difficult for users to avoid being tracked. Additionally, about 90% of the websites involved in the study placed tracking cookies on users' browsers before they were given the chance to opt-out. Utz *et al.* (2019) examined common features of consent dialogs and found that many elements can be used to nudge users to accept tracking. In a field experiment, Godinho de Matos and Adjerid (2021) found that user opt-in for the disclosure of different data types increase if GDPR-compliant consent was used. Using a longitudinal dataset that includes both websites' responses and downstream economic effects, we contribute to the emerging GDPR literature by documenting the evolution over time of content providers' responses to the regulation, and the downstream impact of regulation and websites' responses on the latter's outcomes.

Online advertising and content providers. Research in the online advertising and media literature has investigated the relationship between ad-sponsored business models, content providers' incentives, and the provision of content. Several theoretical studies have argued that when content providers are supported by advertising revenue, they have an incentive to adjust their content to maximize traffic; by so doing, they aim at attracting more advertisers willing to buy ad space on their websites, targeted to specific audiences (Anderson and Gabszewicz, 2006). Empirically, Monic and Feng (2013) investigated changes in the quality of blogs' posts after the implementation of ad-supported business models. They found that the quality of blogs' posts tended to

increase because of ad revenue. Shiller *et al.* (2018) investigated whether the increasing adoption of ad blockers by online users might decrease the quality of online content. The authors used traffic at the website level as a proxy for quality, and found that websites with a high proportion of ad blocking visitors experienced a deterioration in traffic ranking relative to websites with fewer ad blocking visitors. Athey *et al.* (2018) showed how consumer switching – that is, consumers consuming content from multiple websites – affects advertising strategies and, in turn, increases the competition among publishers, leading to an increase in a publisher's incentives to invest in quality content that attracts a greater share of consumers. Once again, to our knowledge, no study has specifically investigated the link between privacy regulation (which may affect the availability of consumer data within the online advertising ecosystem) and downstream economic variables of relevance to content providers, such as their ability to create new content and their success in terms of traffic and social media engagement.

3 Theoretical Framework

The enactment of the GDPR created a privacy regulatory environment with a scope encompassing any entity that operates in the EU or collects the personal data of EU data subjects (GDPR Article 1). This scope is uniquely expansive in both the types of data it covers and its territorial reach - including websites that utilize behaviorally targeted advertising. The GDPR defines 'personal data' to be any data that relates to 'an identified or identifiable natural person' (GDPR Article 4(1)). Due to the nature of how personal data is often used to facilitate behavioral advertising, industry groups and EU governmental bodies have considered that data to be within the scope of the regulation (UK Information Commissioner's Office (2019) International Association of Privacy Professionals (2020)). Additionally, the GDPR is extraterritorial in its scope, applying to any entity that handles the personal data of EU data subjects indepen-

dent of location (GDPR Article 3(2)). Because of this, non-EU websites that utilize behavioral advertising and accept traffic from EU data subjects would be subject to the requirements of the GDPR when interacting with EU visitors.

Article 6 of the GDPR arguably provides the primary mechanism by which the behavioral advertising practices of websites would be impacted. Article 6 establishes six lawful bases for data collection. Any website that collects the personal data of EU data subjects (acting as a 'data controller') must justify that collection under one of the six bases. In the context of online content providers, two of these justifications are generally accepted to apply to advertising practices: 'user consent' and 'legitimate interest' (IAB Europe, 2021).

Under the first justification, data collection can proceed if a user (the visitor) consents to the purpose for which it is being collected (GDPR Article 6(1)(a)). For example, when a user browses a website owned by a publisher, the publisher must request the user's explicit permission to allow cookies to be set on the user's machine and, if so, whether she would also allow tracking cookies by third parties. This differs from the pre-GDPR *de facto* standards for most websites across the world. In absence of regulatory obligations, websites can track users' behaviors by default, often merely informing users that they implicitly consent to tracking by virtue of accessing the website. In addition to requiring that organizations provide a lawful basis for data collection and processing, the GDPR—as noted—establishes steep financial penalties for organizations that do not comply. Both the fine imposed on Google by the French privacy regulatory authority (CNIL) (see Section 1) for lack of valid consent for the personalization of advertising, and advice such as that released by the UK Information Commissioner's Office,[4], suggest that consent has emerged as one of the primary basis for enabling data processing for behavioral advertising under the GDPR.

Under the second possible justification, data controllers (such as websites) can

[4]https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf

collect and process data if it is necessary 'for the purposes of the legitimate interests pursued by the controller' (GDPR Article 6(1)(f)). In the case where websites use legitimate interest to justify data collection, these interests must be communicated to the data subject (GDPR Article 14(2)(b)).

In the following subsections, we first establish how the textual requirements of the GDPR may result in downstream economic effects for publishers. We then break these effects apart into 'ecosystem effects' and 'website-level effects' and describe each in detail. Throughout, we propose three economic expectations that describe how the GDPR may impact the market of online content providers. Finally, we discuss how the evolution of enforcement and compliance since 2018 may weaken the downstream effects of the GDPR by potentially changing how website operators perceive risk resulting from the regulation.

3.1 Downstream Economic Impact

Article 6 of the GDPR includes requirements that—at least on theoretical grounds—may ultimately affect downstream economic variables for websites that generate some or most of their revenues through advertising. First, the restrictions introduced by the GDPR may reduce the extent to which website visitors are tracked. The reduction in tracking may be due to a number of factors and manifest itself in a number of ways, including: through the decision by data controllers (such as websites) to reduce or altogether stop the tracking of EU visitors; through websites' decision to block altogether traffic from EU visitors (in order to avoid potential fines associated with GDPR violations); or through the adoption of consent mechanisms, which present visitors with options concerning the usage of their data, and therefore may increase the portion of visitors who opt-out from tracking and targeting relative to the status quo pre-GDPR. We note that the reduction in tracking may be associated with decisions made by (and which may, in turn, affect) both individual websites and the online ad-

vertising/publishing ecosystem as a whole (for instance, data intermediaries such as advertising networks). In Subsections 3.1.1 and 3.2 below we discuss the subtle differences in the way decisions by different stakeholders may affect websites' downstream outcomes—something our empirical analysis attempts to account for.

A reduction in the ability to track users will, in turn, adversely affect websites' ability to target them with personalized ads when they visit. The availability of personal data makes online advertising more efficient and more accountable, in the sense that advertisers can monitor advertising performance and effectiveness through quantitative metrics (Johnson *et al.*, 2017). This creates a self-reinforcing cycle that increases the efficiency of online advertising. Vice versa, curtailing the collection of personal data and the tracking of online users would negatively affect profiling (Goldfarb and Tucker, 2010).

A reduction in the ability to personalize ads would, in turn, negatively affect their value. Ads that are are tailored to visitors' preferences are more valuable. Personal information increases targeting efficiency (Tucker, 2012). Besides allowing for more granular targeting, online advertising also has significant cost advantages compared to offline advertising (Goldfarb, 2014). Non-targeted impressions may therefore receive lower bid prices in ad auctions Beales (2010). Furthermore, a reduction in the ability to collect and use visitors' data would decrease the number of targeted impressions within ad auctions. Thus, online advertising may become less profitable as whole (Goldfarb and Tucker, 2011), and websites that provide content may receive lower payments from selling advertising space for non-behaviorally targeted impressions (Marotta *et al.*, 2019; Sharma *et al.*, 2019).

In turn, and as a result, overall revenues of content providers may decrease (Lambrecht *et al.*, 2014). Existing work has documented the prevalence of ad-sponsored business models among these websites (Casadesus-Masanell and Zhu, 2013; Goldfarb, 2004; Lambrecht *et al.*, 2014). Both theoretical and empirical works have tied providers'

content quality to advertising revenues (Anderson and Gabszewicz, 2006; Monic and Feng, 2013). Thus, in response to reduced revenue, websites may ultimately not be able to sustain the quantity and quality of output (content) they generated before the regulatory shock Downes (2018).

While we cannot capture changes in revenues directly for a sufficiently large number of websites, we can capture a number of metrics that are correlated with websites' ability to provide quality content. The metrics include variables used before in related literature (Shiller *et al.*, 2018; Gallea and Rohner, 2021; Ferreira *et al.*, 2021), such as the amount of new content URLs generated by online publishers over time, the volume of traffic they receive, and the degree of social media engagement. We discuss such downstream economic variables in Section 4.2.

As noted above, there are subtle but important differences in terms of the process through which choices by various stakeholders (individual websites vs. other agents in the online advertising/publishing ecosystem) may, in theory, disparately affect websites' downstream outcomes. In the rest of this section, we distinguish between "ecosystem" effects and "website-level" effects. Ecosystem and website-level effects may operate both separately and in combination to affect downstream economic outcomes. In section 6 we discuss how our empirical strategies attempt to detect and separate these various possible effects.

3.1.1 Ecosystem Effects

By ecosystem effects we refer to the process through which the responses to the GDPR by the group of all stakeholders in the online advertising/publishing ecosystem (such as individual websites, ad networks, and so forth) may *collectively* affect the aggregate availability of consumer data within that ecosystem, and how such changes in data availability may, in turn, affect the ability of *individual* websites to target their respective visitors with behavioral ads—*regardless of that specific website's own GDPR*

response. As more stakeholders respond to the GDPR by limiting data collection (for instance, by adopting consent mechanisms that facilitate users' opt-outs, or by reducing tracking altogether), less personal information may become available in the ecosystem, and fewer individuals may be precisely profiled for behavioral advertising when they visit any given website in that ecosystem. This, in turn, would affect that website's revenues, and ultimately its content—again, independently of its own distinct GDPR response—through the chain of effects discussed above at the start of Section 3.1.

Put in other terms, ecosystem effects capture downstream economic impacts of the GDPR that arise not directly (through a given website's specific GDPR response), but indirectly (through changes in the overall availability of consumer data across the entire advertising and publishing ecosystem). For instance, following the enactment of the GDPR, the profiles of existing users within ad-tech companies or ad platforms may become less accurate or less up to date, as data becomes more sparse. In turn, the broader advertising ecosystem may see lower bids in ad auctions for users with less detailed profiles (Goldfarb and Tucker, 2011). In addition, the amount of behavioral data associated with a random visitor to a specific website may, on average, decrease. In both cases, we should expect a reduction of websites' advertising revenues.

It stands to reason that the revenue loss due to this ecosystem effect—and therefore the resulting reduction in websites' ability to provide quality content—should not be expected to be the same for EU websites relative to US websites. First, while EU-based data controllers (such as websites and advertising networks) are required to comply with GDPR rules for *all* users and visitors, this is not the case for data controllers based outside the EU, which may choose to apply GDPR-compliant practices only to the share of visitors originating from the EU. We expect this to reduce the overall availability of user data within the EU-based advertising ecosystem more intensely than any reduction within the US-based advertising ecosystem. In addition, we expect EU-based websites to receive a higher percentage of traffic from EU-based visitors (that

is, visitors from a EU IP address) than US-based websites; therefore, we expect a higher proportion of traffic to EU-based websites to be less precisely trackable—and therefore less precisely targetable—than traffic to US-based websites. In short, we expect the magnitude of ecosystem effects on website-level downstream economic variable to be moderated both by where websites are based and by the share of each website's traffic that originates from the EU, independently of a given website's GDPR response. Specifically, we expect EU websites, on average, to be more negatively affected by ecosystem effects than US websites, as the GDPR should more significantly reduce tracking and data availability across the EU-based data ecosystem, thus affecting all websites in it regardless of their response. In addition, we expect websites with higher percentage of EU-based traffic to be more affected than websites with lower percentage of US traffic, at parity of type of GDPR response by the website. We expect this to be the case regardless of the geographical location of the website: in other words, if both a EU-based website and a US-based website responded to the enactment of the GDPR by reducing tracking and targeting of EU visitors, we would expect the overall impact on revenues (and thus, ultimately, on downstream economic outcomes) of their response to be larger for the EU-based website, because (in this example) the site with a higher percentage of EU traffic would be curtailing tracking and targeting of a higher percentage of its traffic, relative to the website with a lower percentage of EU traffic.

We test the validity of our assumptions, and attempt to disentangle these ecosystem effects, in Section 6. In Section 4 we discuss nuances associated with identifying EU and US websites based on domain, traffic patterns, and headquarter location.

3.2 Website-Level Effects

By website-level effects we refer, instead, to the process through which responses to the GDPR by a given website may *individually* affect that website's ability to collect visitors' data and/or use it for behavioral advertising. At the website-level, the per-

sonal information collected during each visit enables both the tracking of visitors (this information may be shared with and sold to other players in the the rest of the ad-tech ecosystem) and the targeting of ads to visitors on that website (the targeting, itself, may rely on a combination of user data coming from both the website and its partners in the ecosystem). Thus, websites' responses that limit data collection or usage (for instance, the adoption of consent mechanisms that allow visitors to opt out of tracking or targeting) may affect that website's revenues (Sharma *et al.*, 2019), and ultimately its ability to provide quality content.

Different websites' response to the GDPR may have heterogeneous repercussions on the chain of effects discussed at the start of Section 3.1, and thus may disparately affect downstream economic variables. We define five categories of website-level responses, and discuss them in the rest of this subsection, starting with the most aggressive and ending with, arguably, the most lax.

First, faced with the compliance burden imposed by the GDPR, websites—especially those based outside the EU—have the option of exiting the data and advertising market altogether, by blocking all traffic originating from the EU. In our analysis we refer to this response as *"Blocks EU"*. This response would arguably result in the strongest negative effects on downstream economic variables for the website, as cutting off EU visitors would directly curtail potential future advertising revenue. This option may be attractive for websites based outside the EU which received only a small share of traffic from the EU prior to May 2018. The exit of these websites from the EU market may redirect their former users to other websites (if substitutes exist), but would likely not result in negative ecosystem effects on the tracking ability or targeting accuracy of other websites.

Second, websites may respond to the GDPR by unilaterally reducing or halting the tracking and targeting of EU visitors. We expect this response to also negatively affect a website's advertising revenues, although arguably not as intensely as blocking

EU visitors from accessing the website. We will refer to this response as *"Stops EU Tracking"*.

Third, websites may display consent mechanisms to vistors for the purpose of obtaining user consent to engage in data collection and data usage. Such implementation of consent mechanisms may, too, diminish websites' ability to collect personal information—albeit arguably to a lesser extent than the unilateral curtailing of tracking and targeting by a website. Visitors to websites that implement consent dialogs may, for instance, not consent to tracking for the purposes of targeted advertising. From the perspective of websites, these visitors would no longer be linkable with interest profiles used for targeting ads. These effects can vary in magnitude depending on the specific manner in which websites choose to implement consent dialogs, including interface features and the possible deployment of dark patterns (see Section 3.3). Websites which implement consent dialogs that make it easier for users to deny consent for tracking (such as dialogs that only require a single-step to reject tracking) may experience stronger negative effects on tracking and targeting ability compared to websites that make denying consent for tracking more difficult (such as websites that implement consent dialogs which require multiple steps to reject tracking).

Fourth, websites may attempt to minimize the impact of having to implement a consent mechanisms by instituting 'cookie walls'. These dialogs force users to first consent to tracking before allowing them to view content. By forcing consent, these websites may not see a decrease in their ability to track visitors. However, visitors that do not wish to be tracked may be turned away. While the legality of this response under the GDPR is unclear ((UK Information Commissioner's Office, 2019), (Autoriteit Persoonsgegevens, 2019)), we observe multiple websites within our data which use cookie walls (see Section 5). Given the disparate effect different ways of eliciting consent may have on downstream effects, in our analysis we distinguish between *"Consent Mechanism"* for websites that allow users to opt-out of tracking without limiting access to

the website, and *"Cookie Wall"* for those that force consent before giving users access to the website.

Fifth, websites may not take direct actions in response to the GDPR. This category is broad. Some websites may elect to not curtail tracking nor implement consent mechanisms, but rather invoke 'legitimate interest' (see Section 3) to justify continuing their present data collection and usage practices.[5] Other websites may simply continue to comply with older EU privacy directives, merely displaying 'cookie notices' which often appear as banners at the bottom of websites.[6] Functionally, the effects on tracking and targeting for the websites invoking GDPR legitimate interest and websites not even bothering doing so may be the quite similar. Either way, these websites do not engage in compliance changes that are likely to affect the trackability of their visitors (if anything, they may even benefit from the reduced tracking ability of other websites, as a decrease in tracked advertising inventory may drive up advertisers willingness to pay). Theoretically, they may experience a reduction in traffic from privacy-conscious and aware visitors who dislike the imposition of tracking, without consent, based on the legitimate interest rationale. In practice, we expect this category of websites to experience the mildest effect on revenues and thus on downstream variables. For our analysis, as all these responses are unlikely to have a significant impact we group them together and refer to them as *"No Response or Legitimate Interest"*.

To summarize, we expect the magnitude of website-level effects on websites' revenues, and thus on downstream economic variables, to vary based on how websites choose to respond. Some responses may lead to stronger negative impacts on tracking

[5]The legality of this justification for tracking is contested, and the compliance risk potentially high. As early as 2019, the UK Information Commissioner's Office published an opinion stating that legitimate interest cannot be used as a legal basis for data collection in the context of behavioral advertising (UK Information Commissioner's Office, 2019). This has grown into a consensus among regulators and industry over time. Earlier this year, IAB Europe published guidance stating that legitimate interest cannot be used as a basis for setting tracking cookies IAB Europe (2021).

[6]The banners inform users of the presence of cookies on a website and are distinct from other privacy notices in that they do not ask for consent prior to tracking or notify users of legitimate interest claims).

and targeting, which may result in stronger negative impacts on downstream economic variables. We expect that websites which respond to the GDPR by blocking EU users or curtailing tracking and targeting to experience the strongest negative downstream effects, followed by websites implementing consent mechanisms with explicit options to reject tracking with a single-step, followed by websites implementing consent mechanisms designed to nudge visitors towards opt-in (for instance, those that requires multiple steps to reject tracking), followed by websites that do not actively respond to GDPR and/or claim legitimate interest. Of course, how a website chooses to respond will be influenced by factors including the percentage of EU traffic they receive. For instance, websites that choose to block EU visitors are likely those that only receive a negligible share of traffic from EU visitors. We account for this endogeneity in our empirical analysis (Section 6).

3.3 The Evolution of Enforcement and Changes in Downstream Economic Impacts over Time

In the years since the GDPR came into effect, both compliance and enforcement have been slow to ramp up (Massé, 2021). Enforcement and compliance have evolved over time. This evolution should be reflected in changes over time in both websites' responses, and associated downstream economic impacts. We rely on a longitudinal dataset spanning nearly two years in order to capture potential differences between short-term and long-term effects of the GDPR.

The slow ramp-up in part reflects uncertainty on the part of EU regulators and industry regarding what it means to be GDPR compliant. Through the release of guidelines and enforcement actions by member states' Data Protection Authorities (DPAs), a working definition of compliance may be starting to emerge. However, the delays in enforcement actions and incentives among publishers to preserve the status-quo of tracking visitors may blunt the downstream effects of the GDPR on content.

While the number of fines issued by DPAs has increased recently, many still face a backlog of complaints and investigations (Massé, 2021). Only within the past few years have some DPAs began to issue guidance on cookies and enforcement decisions for non-compliant data controllers and processors. In some cases, these actions may have impacted the risk assessments by content publishers deciding whether and how to comply with the requirements of the GDPR. For example, guidance on how websites should comply with GDPR cookie rules were only finalized by the French DPA (CNIL) in September 2020 (Commission nationale de l'informatique et des libertés, 2020). Prior to this, the CNIL enforced its pre-GDPR cookies guidelines from 2013. Following the finalization of these guidelines, the CNIL granted websites a six month 'grace period' to comply. Because of this, cookie rules taking into account the requirements of the GDPR only started to be enforced in April 2021 (Commission nationale de l'informatique et des libertés, 2021).

In response to the uncertainty surrounding enforcement, websites have adopted a diverse set of compliance strategies and behaviors. Many websites introduced or updated privacy notices immediately before and after the GDPR enforcement date to include consent options. Early studies indicated that these implementations vary in terms of functionality and ease of use. In a number of cases, early consent dialogs did not change the tracking behavior of websites if users chose to reject tracking (Sakamoto and Matsunaga, 2019). The increase in consent dialogs also saw an increase in 'dark patterns' or user interfaces designed to make it difficult for users to reject tracking (Nouwens *et al.*, 2020).

One of the largest impacts on the design and usage of consent dialogs may have been the introduction of Consent Management Platforms (CMPs) which help websites integrate consent into their advertising stack. In a recent study, Hils *et al.* (2020) track the adoption of CMPs over time following the GDPR across a broad range of websites. The authors find that adoption of these consent platforms is most prominent among

mid-ranking websites. They additionally find that many of the consent dialog interfaces vary and that many ad vendors continued to justify data collection for tracking using legitimate interest. Furthermore, the authors tested different variations of CMP consent dialogs, finding that those which required more steps to reject cookies took longer for users to navigate and increased the percentage of users choosing to accept cookies.

In addition to designing consent dialogs that encourage visitors to consent to tracking, the ad tech industry has sought to identify other means to enable visitor tracking. One emerging method involves using first-party cookies that are set directly by publishers. Unlike third-party cookies, users may more readily consent to first-party cookies. Several publishers have started to use first-party cookies for tracking, however no single method for using first-party data to facilitate targeted advertising has yet to emerge (Kaye, 2021)(O'Reilly, 2020).

Given the slow movement of industry and regulators towards a consensus on compliance (Lancieri, 2021), many of the downstream economic impacts on publishers may not be visible in the short-term even if they are present. We examine the downstream effects on publishers in the 18 months following the enforcement date of the GDPR to gain a longer-term perspective on the regulation.

4 Study Design

The study design relies on a sampling strategy for websites whose data we planned to collect (Section 4.1) and a set of metrics periodically captured for each of those websites (Section 4.2).

4.1 Websites Sample Selection

We constructed a large longitudinal panel of websites located in the EU and in the US. While our actual panel includes different categories of websites, in this manuscript we

focus exclusively on its subset of content providers, such as news websites and online magazines, due to their reliance on online tracking and behavioral targeted advertising as the main source of revenue.

To examine the impact of the GDPR on both major and minor websites (and specifically, in the context of this manuscript, on content providers), the panel includes both top ranked and long-tail domains from the US and from several EU countries (Germany, France, UK, Italy, Spain, and the Netherlands).[7] We used 2018 Alexa data to identify top ranked websites. At the time, Alexa data provided the top 500 websites from various geographical areas (including Global, Germany, France, Italy, Spain, Netherlands, and USA.) and five content categories (News, Sports, Society, Health, and Games).[8]

Alexa's top 500 websites by country correspond to the websites most *visited* by users in that country (rather than the most popular websites that are *based* in that country). To include the top websites based in each of our areas of interest (EU and US), we used Alexa's global top 1 million websites to complement the dataset with the top 500 websites in various top-level websites, such as *.au, .de, .fr, .uk, .it, .es, .nl, .com, .net* and *.us.*.

We complemented highly ranked websites with a random sample of websites ranked between 200,000 and 1 million. Specifically, we included in the panel 500 random websites for each 100k websites ranking interval, i.e. 500 websites ranked between 200k and 300k, 500 websites ranked between 300k and 400k, and so on until reaching 1 million.

The resulting sample included 11,254 websites. We eliminated websites that only got a minor fraction of their visitors from EU countries or the US, despite the fact

[7]The country of a website was determined by the location of its headquarters as reported by SimilarWeb. When this information was not available, we defined the country of a website by using the website's top-level domain country of origin. In the case where neither the country nor the top level domain are available, we assigned the website to the country where most visitors originated from.

[8]In September 17, 2020 Alexa discontinued its categories ranking. This does not affect our data, as it happened well after our sample was created.

that they were among the most popular websites in one of our countries of interest.[9] Finally, we scrutinized the content categories of the remainder websites. We noticed that the content categories provided by Alexa were often inconsistent. Thus, we obtained categories information using SimilarWeb,[10] which in our experience provided a more robust categorization. We excluded from the sample all websites categorized as providing adult content, or not assigned to any category.

The resulting set contains 5,474 websites. Of them, 909 were in the News and Media category. This sample is the focus of our analysis.

4.2 Data Collection

For each News and Media website in our sample, we collected two sets of metrics at regular intervals. The first set of metrics consisted of data we directly mined from the websites in the sample using a web privacy measurement framework (OpenWPM). We refer to these variables as "technical variables" (see Section 4.2.1). The goal of collecting these metrics was to capture websites' response to the GDPR, and in particular to determine whether and how websites' behavior (including tracking, privacy, and advertising choices) changed after the implementation of the GDPR. Technical variables were captured by browsing each website simultaneously from two different visitor IP addresses, one located in Europe (France) and one in the US. This design allows us to compare, before and after the enactment of the GDPR, whether and how websites adapted their data collection behavior according to the geographical location of a visitor.

The second set of metrics were obtained from third parties repositories. With

[9]For example, the Russian shopping website *avito.ru* was the 52nd most visited website in The Netherlands in May 2019. However, visitors from The Netherlands account for less than 2% of all avito.ru's visitors, while visitors from Russia account for roughly 85%. Therefore, although the website is popular in at least one EU country, it would be unreasonable to assume it will significantly change its behavior due to a European regulation considering that it is a Russian website that gets most of its visitors from Russia.

[10]See `https://www.similarweb.com/`.

these metrics, we aimed to measure changes in the quantity of content offered by the websites in the sample and in user engagement with such content (a proxy for its quality). We refer to these variables as "downstream economic variables" (see Section 4.2.2). These metrics do not change as function of the country of the visitor. However, we do expect to find differences depending on the location of registration of the website, as websites registered in different locations (EU vs US) should be affected differently by the GDPR.

The data collected spans a period of time of at least 19 months for technical variable metrics (from April 2018 to November 2019), and 31 months for downstream economic variable metrics (from April 2017 to November 2019). Our empirical analysis attempts to tie changes in the downstream metrics to websites' and ecosystem's responses to the GDPR.

4.2.1 Technical Variables

We rely on OpenWPM, a web privacy measurement framework (Englehardt and Narayanan, 2016), to simulate user browsing and investigate how websites interact with their visitors, thus capturing website-level responses to the GDPR. The framework is implemented within an instrumented web browser that automates the process of visiting a set of websites and records a series of variables. We refer to each round of visits to all websites as a"wave" of data collection. During each wave, we visited each website twice, once with an US visitor address and once with an EU visitor address.

The metrics collected via OpenWPM cover different websites' behaviors (including tracking, privacy, and advertising choices). In addition to the raw metrics collected via OpenWPM, we further processed the data to construct additional metrics related to advertising shown on the websites in the panel. Using scripts included in popular ad blockers, we flagged advertising content within the HTML content we extracted via

OpenWPM.[11]

The metrics collected via OpenWPM, or constructed by us through a variety of tools, are summarized below.

Cookies: Cookies are small files stored on visitors' browsers and often embedded on websites to provide additional functionality. Cookies are extensively used for advertising purposes—for example, to store information on the websites or products visited by a user. Our data collection focused on two types of cookies: 1st party and 3rd party cookies. The variable *1st Party Cookies* measures the cookies which are set by the website being browsed. The variable *3rd Party Cookies* represents cookies that are set by entities other than the original website, and that could be used to track users' behavior across different websites in order to construct users' profiles aimed, in part, at improving behaviorally targeted ads.

Advertising content length: To analyze the volume of advertising displayed to visitors when browsing websites in our panel, we captured the length (in bytes) of certain types of websites' HTML content.Of particular interest to our analysis is the length of advertising content, which allows us to get information about the amount of ad content displayed on a given website's homepage. The variable *Ads Length (KB)* provides information on the size in kilobytes of the quantity of advertising content on a website. It is constructed by measuring the length of the content that is identified as advertising by *Adblock Easylist*[12].

[11]An ad blocker is a small piece of software or module incorporated into a user's browser (Add-on) that prevents the display of banners and other advertising formats. Ad-blockers filter advertisements by recognizing the advertising tags of the main ad servers and advertising networks. We cross-referenced the data we collected from OpenWPM with these filtering lists (blocklists). (List inside ad-blocker add-on to block unwanted content like advertising.) In particular, we rely on two blocklists: Adblock Plus (`https://adblockplus.org/fr/subscriptions`. Last retrieved, February 2020) and Disconnect (a free extension for the web browser responsible for blocking trackers from web pages that the user visits: `https://disconnect.me/`), which establishes identification and classification rules for advertising and tracking entities.

[12]AdBlock Easylist consists of a set of rules used by AdBlockers to detect and hide elements that correspond to advertising. We re-purpose these rules to identify and measured the length of advertising

Websites user interaction: We capture and classify visual elements of websites which may indicate a website's response to the GDPR using a combination of automated HTTP response analysis[13] and manual coding of screenshots captured by OpenWPM. These elements include buttons (such as buttons to accept cookies) and user-facing messaging (such as information about cookies or website availability). We categorize these visual elements into four groups, which we then use in our analysis to help differentiate website response. First, US websites which decided to get out of the European market by simply blocking EU consumers (visitors) often indicate this by displaying a static page to EU visitors informing them that the website is unavailable (see Figure 9). Second, websites that request the user's explicit permission to allow cookies to be set on the user's machine do so by implementing a consent mechanism. We consider a website to have implemented a consent mechanism if it contains at least one "accept" button (for example Figure 10a and Figure 10b are considered as consent mechanisms although in Figure 10b there is no reject button). Third, we distinguish the consent mechanisms with the cookie wall by the fact that the cookie walls prevent visitors from viewing content and do not provide a means (through buttons or links) to reject tracking (see figure 11. Finally, we classify cookie notices which inform users about cookies but do not provide means to accept or reject them as 'cookie banners' (see Figure 12). In other words, there is no direct action from the user.

4.2.2 Downstream Economic Variables

Content-related metrics were collected from third parties' repositories to measure downstream changes in quantity of content generated by websites in the panel, and changes in user engagement with such content (a proxy for content quality).

To measure content quantity, we used the *Global Database of Events, Language,*

instead of hiding it. The list is available at: `easylist.to`

[13]HTTP responses measure all the information exchanged between the browser and the websites that are visited by a user. By capturing all HTTP responses, we can identify the number of responses linked to a pop-up window such as consent mechanism or cookies banner.

and Tone (GDELT) to capture the number of new URLs of content published by each website in our sample in the week surrounding each observation from OpenWPM (three days before and after each OpenWPM observation).[14] GDELT gathers and provides metadata for articles from news and media websites going back to 2015 from both domestic (US) and international sources. The database provides metadata including the URL, publication date, and publisher website for each article, and has been used in studies which examine global events (Gallea and Rohner, 2021; Ferreira *et al.*, 2021). Because we visit each website multiple times to construct our longitudinal data set, we collect multiple observations of the new URL counts for each website over time.

As a proxy for quality, we measure visitors' engagement using websites' traffic metrics (Page Views Per User, Rank) and social media reactions (Facebook Reactions). The underlying premise is that, were the quality of the content provided by the website to decrease, users may try to substitute for other content and, therefore, we should observe a decrease in the number of visits to a given website. This is similar to the strategy used by Shiller *et al.* (2018) to measure websites' quality.

Specifically, we use Alexa's *Rank* as a measure of the number of users visiting a website. It is calculated by looking at the estimated average daily unique visitors and page views for a given website. We also use Alexa's *Page Views Per User*, which represents the number of pages viewed per user on a website. Finally, we augment the traffic variables with data on the number of social media "reactions" related to the content published on the websites in our sample. We use data from Facebook Graph API, in line with Cagé *et al.* (2015) methodology, who used the same metric as a proxy of quality for online news websites. Specifically, for each new URL of content posted by each website in our sample during the week surrounding the data collection in each wave (as retrieved via GDELT), we collect the number of reactions on the Facebook platform and calculate the average number of *Facebook Reactions* across all new URLs

[14] `gdeltproject.org`

by website/wave. We call this average the *FB Average Reaction*. Such reactions can be used to measure users' engagement with a piece of content, and can be therefore interpreted as a proxy for content quality. Table 1 presents the descriptive statistics of the technical variables and the downstream economics variables.

Table 1: Descriptive Statistics - Overall Sample

	Mean	Std. Dev.	Min	Max	N
		Technical variables			
Tracking:					
1st Party Cookies EU Visitor	11.058	8.008	0.0	52.0	11114
3rd Party Cookies US Visitor	33.269	37.620	0.0	272.0	11114
1st Party Cookies US Visitor	12.942	8.288	0.0	50.0	11175
3rd Party Cookies EU Visitor	53.892	50.060	0.0	351.0	11175
Advertising:					
Ads Length (KB) EU Visitor	576.564	1279.351	0.0	111046.5	11107
Ads Length (KB) US Visitor	695.740	932.966	0.0	28571.3	11172
Website level responses:					
Blocking EU Visitor	0.023	0.151	0.0	1.0	21798
Stop EU Tracking	0.118	0.322	0.0	1.0	11114
Consent Mechanism EU Visitor	0.133	0.340	0.0	1.0	21798
Cookie Wall EU Visitor	0.017	0.130	0.0	1.0	21798
Cookie Banner EU Visitor	0.101	0.302	0.0	1.0	21798
Website share Visitor:					
Share of EU Visitors	0.430	0.420	0.0	1.0	21798
Share of US Visitors	0.395	0.403	0.0	1.0	21798
		Downstream variables			
Dependant variables:					
Log GDELT URLs	5.014	1.705	0.0	9.6	17588
Page Views Per User	2.032	0.953	0.6	19.1	21797
Rank	65063.040	100954.740	0.0	1832762.9	21797
FB Average Reaction	110.463	466.708	0.0	12476.9	17588
Observations	21798				

Notes: This table presents the descriptive statistics for the overall sample (i.e. including all observations from all websites and waves).

5 Descriptive Statistics

In section 3.2 we explained how websites could adopt different responses, and identified five types: Blocks EU; Stops EU Tracking; Consent Mechanism; Cookie Wall; and No Response or Legitimate Interest. The overall effect of the GDPR on the extent of tracking is thus not obvious, as it will depend on the number of websites that adopt

each response, and the impact of each response on tracking. In this section, we leverage the technical variables we collected to describe to what extent tracking changed in EU vs US websites. Our intent is not yet to make any causal claim regarding the influence of GDPR in these changes. Instead, we present these empirical patterns to describe the behavior of the variables we collect before GDPR was enacted, and both in the short and the long-term after the GDPR was enacted.

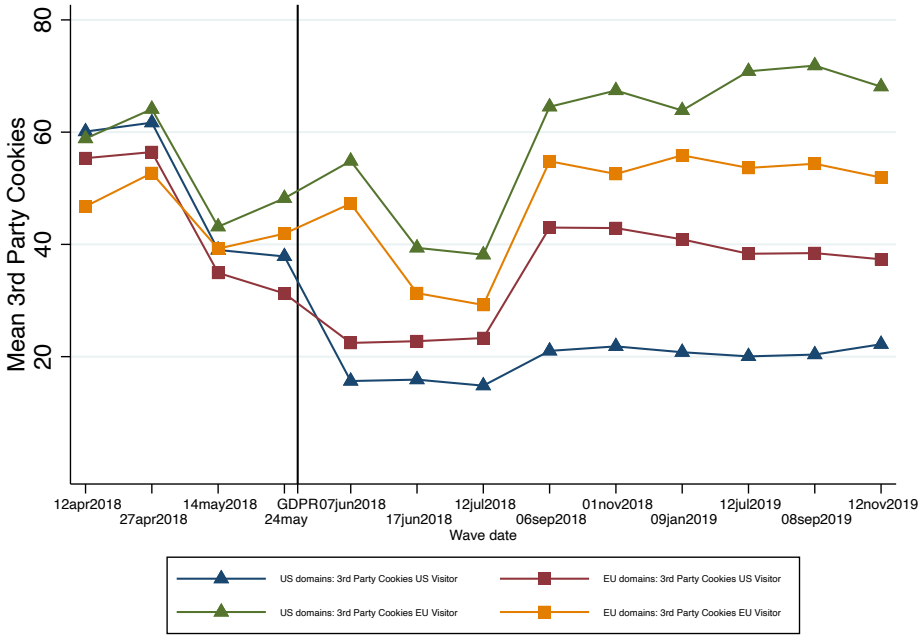
5.1 Changes in Tracking and Advertising Traffic

We first present general trends on the evolution of the internet traffic related to tracking and advertising that results from visiting the websites in our sample before and after the GDPR became effective. We contrast EU vs US based websites, and how the results change if the websites are visited from EU or US based IPs.

We first focus on 3rd party cookies, which are typically used to track users across websites. Our data collection strategy allows us to distinguish and observe four scenarios: how websites based in the US (US domains) treat visitors that originate from the EU (EU visitor) or from the US (US visitor); and how websites based in the EU (EU domains) treat EU visitors and US visitors. Figure 1 shows how, before the GDPR, the number of 3rd party cookies used by EU and US domains were similar for both EU and US visitors. At most, the number of 3rd party cookies set by EU domains was slightly lower, especially for EU visitors. Even before the GDPR came into effect, we observed a drop in the number of 3rd party cookies being used in EU/US domains for both EU/US visitors. The sharpest drop we observed happens in US domains for EU visitors, followed by EU domains for EU visitors. However, these drops seem to be short lived, as we observed a rebound in the number of 3rd party cookies set by websites roughly 3 months after the GDPR became effective. Notably, the rebound is not the same for all websites and visitors. For EU visitors, US domains continue to set a much lower number of 3rd party cookies than before the GDPR. In the case of EU

domains visited from the EU, the number of 3rd party cookies rebounds, but stays at a lower level than before the GDPR. For EU and US domains browsed by US visitors, we observe that, in the long-term, the number of 3rd party cookies goes back to the pre-GDPR level or even higher. It is also apparent that EU domains set fewer cookies than US domains (except in the case where US domains are visited from the EU).

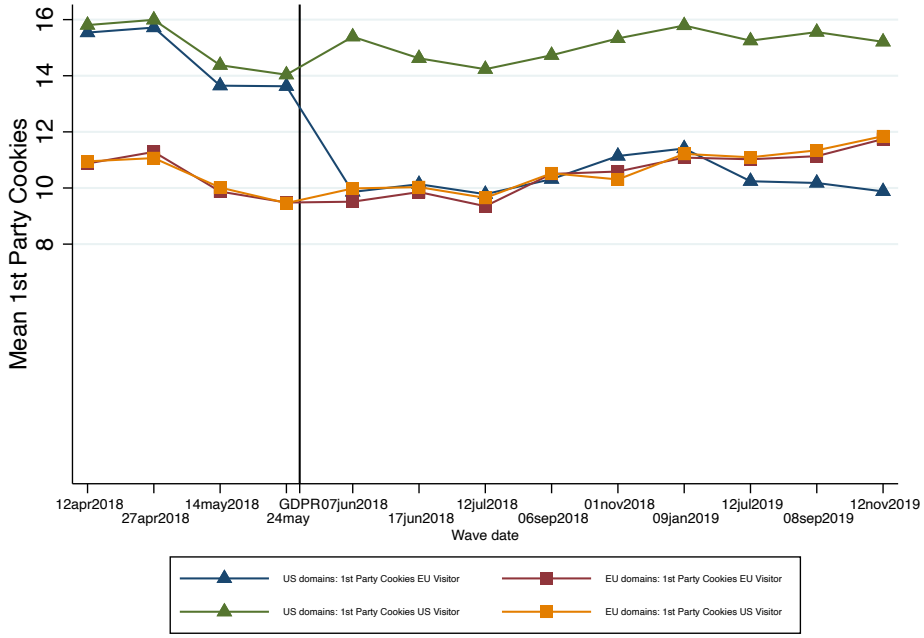
Fig. 1 *3rd Party Cookies Set by EU/US Domains for EU/US Visitors*



After documenting an apparent persistent decline in the number of 3rd party cookies set by websites for EU visitors, we examined if the number of 1st party cookies used by websites had also changed. While 3rd party cookies are typically used to track users across websites, 1st party cookies are typically related to particular websites functionalities. For example, a website may use 1st party cookies to remember visitors' login information, products they have browsed, or news articles they have read. However, since 1st party cookies can also be used for advertising purposes, we are interested in examining whether the use of 3rd party cookies is being replaced by 1st party cookies. Figure 2 suggests that the number of 1st party cookies set by websites remains

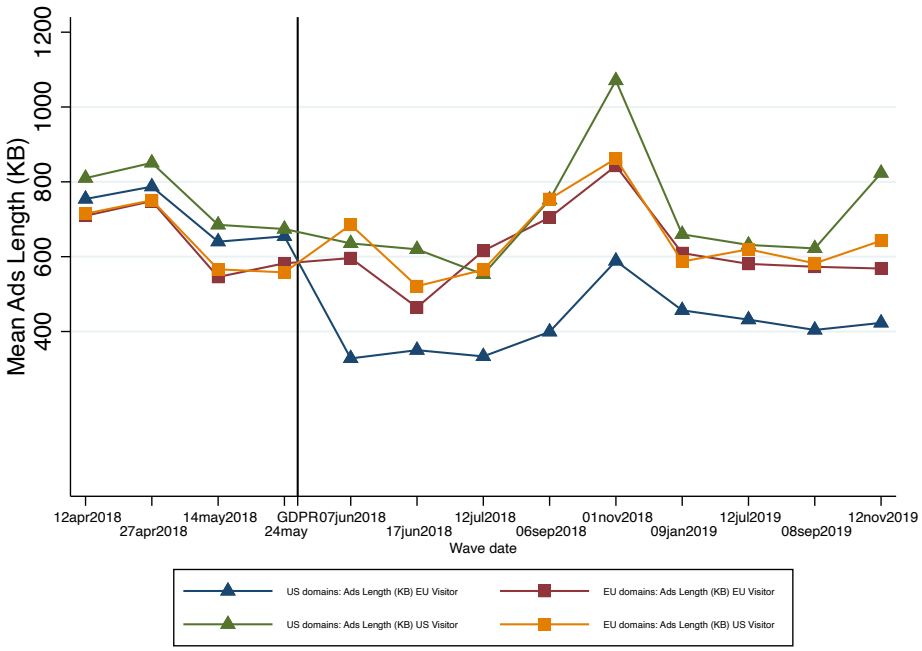
unchanged over time, except for the case of US websites when visited from the EU, for which we observe a persistent drop after the GDPR. From the figure it is also clear that EU websites seem to set, on average, fewer 1st party cookies than US websites.

Fig. 2 *1st Party Cookies Set by EU/US Domains for EU/US Visitors*



While 3rd party cookies are typically used by advertising technology firms to track users across websites (frequently with the goal of inferring their interests and the products they have browsed in order to be able to deliver behaviorally targeted ads), 3rd party cookies can also be used for other purposes. To get a more precise measure of the amount of data being sent to websites' visitors for advertising purposes, and the reliance of websites in our sample on advertising, we explore how advertising length (as defined in section 4.2.1) evolved over time. In Figure 3, we observe that GDPR caused a decrease in advertising length for EU websites. For US websites, the response is more nuanced. While over the long-term advertising length for US visitors seems to return to pre-GDPR levels, it remains at a much lower level for EU visitors.

Fig. 3 *Advertising Length EU/US Domains for EU/US Visitors*



Looking at the evolution of technical variables related to tracking and advertising reveals that, in fact, there is a response to the enactment of the GDPR, but the response is heterogeneous. EU and US websites respond differently. Moreover, their responses are influenced by the location of the user. In the next subsection we study the prevalence of the 5 stylized responses to GDPR we identified in section 3.2 over time and characterize the different websites using them.

5.2 Website-level Responses to GDPR

Table 2 shows descriptive statistics of pre-GDPR website level characteristics, for the different types of website-level responses we identified in section 3.2. In the following subsections, we analyze the features of websites in each response category.

Table 2: US and EU Websites Characteristics Before GDPR Based on Their Most Prevalent Response to GDPR

	EU Websites				US Websites				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Stops EU Tracking	Consent Mechanism	Cookie Wall	No Response or Legitimate Interest	Blocks EU	Stops EU Tracking	Consent Mechanism	Cookie Wall	Nothing or Legitimate Interest
	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd	mean/sd
Websites characteristics									
Rank	113408.73	47015.65	82337.21	31901.61	114518.06	89405.21	39871.92	1979.97	75183.98
	(155539.62)	(95122.03)	(65529.04)	(51758.41)	(108742.59)	(82685.77)	(90162.72)	(2493.84)	(105282.56)
Share of EU Visitors	0.90	0.78	0.96	0.80	0.02	0.03	0.07	0.04	0.04
	(0.25)	(0.24)	(0.04)	(0.23)	(0.01)	(0.05)	(0.11)	(0.01)	(0.04)
Share of US Visitors	0.01	0.04	0.00	0.03	0.90	0.84	0.66	0.70	0.77
	(0.02)	(0.07)	(0.00)	(0.10)	(0.10)	(0.19)	(0.22)	(0.09)	(0.20)
Ads Length (KB) EU Visitor	85.20	768.36	176.47	642.96	1105.17	631.17	726.42	903.43	652.41
	(228.42)	(782.98)	(322.63)	(651.78)	(714.43)	(510.68)	(575.59)	(661.26)	(642.79)
Cookie Banner EU Visitor	0.06	0.10	0.03	0.11	0.00	0.01	0.01	0.02	0.01
	(0.24)	(0.30)	(0.16)	(0.31)	(0.00)	(0.12)	(0.09)	(0.15)	(0.08)
Privacy									
3rd Party Cookies EU Visitor	8.02	52.07	11.52	45.77	58.26	59.86	60.96	86.78	41.70
	(24.97)	(46.79)	(23.11)	(45.11)	(30.93)	(39.86)	(46.79)	(60.49)	(40.70)
1st Party Cookies EU Visitor	3.25	11.51	5.61	10.55	10.40	17.43	15.58	15.61	14.49
	(2.95)	(6.55)	(4.25)	(6.69)	(5.68)	(8.63)	(9.41)	(3.26)	(9.07)
Obs.	225	3,060	390	3,283	688	1,050	1,304	90	3,519
Unique websites	16	204	26	219	46	70	87	6	235

Notes: This table presents the websites pre-GDPR characteristics for the different type of website-level responses. Most prevalent response refers to the response that we observe most frequently for each website across all the waves of data collection.

5.2.1 Blocks EU Visitors

A number of websites decided to exit the EU market by blocking EU visitors from accessing their websites altogether. Websites in our sample that implement such response are all US based and, in general, the overwhelmingly majority of their visitors are US visitors. This type of response was quickly implemented after the GDPR, and the share of US websites using this strategy remains fairly constant across all our waves of data.

Before the GDPR, US websites blocking EU visitors received, on average, 90% of their visits from the US, while US websites not blocking EU visitors received only 77% of their traffic from US visitors. In terms of traffic, they rank lower than other websites (and therefore receive less traffic). In terms of reliance on advertising, they seem to rely on advertising to a greater extent than other websites. In summary, our data suggest that the US websites that block EU visitors are smaller content providers (in terms or traffic); they don't receive a lot of traffic from EU sources but they do rely heavily on

advertising. Taking into consideration the characteristics of this group of websites, their response seems sensible: the gain they would obtain from the few EU visitors is very small, compared to the potential costs of compliance to the GDPR requirements and/or potential liability. As such, it is very plausible this group of websites decides to stop having any EU visitor rather than having to incur the cost of implementing the GDPR's requirements and being exposed to the risk of potential fines due to non-compliance or data breaches.

5.2.2 Stops EU Tracking

Instead of completely blocking EU visitors, websites may instead choose to stop tracking EU visitors after the enactment of the GDPR. To detect this response, we identify all websites that have decreased their number of 3rd party cookies to zero (note that we also include in this group websites that before the GDPR were not using 3rd party cookies and continue to not do so after the GDPR). In our sample, a total of 70 out of 443 US websites chose to stop tracking EU visitors, and only 16 out of 464 EU websites did so. Looking at Table 2, we notice that US websites that decide to Stop EU Tracking have a larger proportion of EU visitors than US websites that decide to block EU visitors, but seem to rely less on advertising. EU websites that decide to not track have a considerable share of EU visitors, but their average advertising length is much lower than EU websites that respond in other ways. These patterns suggest that "Stops EU Tracking" is a plausible response for websites that seem not to rely much on advertising.

5.2.3 Consent Mechanism

Before the GDPR, almost no US websites implemented consent mechanisms (where visitors were given an option to reject tracking cookies), while about 26.04% of EU websites did. The presence of consent mechanisms increased sharply right before the

GDPR became effective, and continued to rise until reaching a stable level with about 68.49% of websites in our sample using them. The websites that chose to implement a consent mechanism were highly ranked (thus, have more traffic), compared to the other groups; they also have a sizeable share of EU visitors (both in the case of US and EU websites); and have a greater reliance on advertising, as suggested by the average advertisements lengths on their websites.

5.2.4 Cookie Wall

The category cookie wall includes websites that implement consent dialogs which, in fact, force users to consent to tracking in order for the user to be able to access the website's content. This strategy can have mixed effects, as if visitors consent to the tracking, the website may not experience any real change in the ability to track visitors. If instead visitors do not wish to be tracked, they may leave the website. EU websites that fall in this category have a large proportion of EU visitors, but tend to be smaller websites (in terms of ranking) and do not rely as much on advertising. The US websites that fall in this category tend, instead, to be highly ranked (they have more overall traffic) and and have a great reliance on advertising, as suggested by the average advertisements lengths on their websites.

5.2.5 No Response or Legitimate Interest

Finally, the last category includes websites that claim legitimate interest, and therefore continue to collect/use data as before the GDPR; as well as websites that decide to not respond. We combine these two types of responses together because we believe that websites in this category will probably not see any effect on our downstream variables of interest: if the websites do not actively change anything following the GDPR (either because they are legitimately doing so or because decide to not act), we should see no systematic changes in the corresponding outcome variables.

The fraction of websites that fall inside this category is larger than what we expected. In terms of characteristics, EU websites that fall in this category have a higher rank, compared to the other websites; and still rely heavily on advertising. US websites that fall into this category also rely considerably on advertising but they rank lower compared to US websites that decide to implement a consent mechanism; and they rank higher than US websites that decide to not track or completely block EU visitors.

5.3 Evolution of Website-level Responses over Time

Website responses to regulation may evolve over time. Figure 4 and Figure 5 summarize the evolution of the response over time for the US website and the EU websites using Sankey diagrams. To make the diagrams readable, we divide the post-GDPR period into three time windows. The first is from May 24, 2018 to July 2018; the second from September 2019 to January 2019; and the last from July 2019 to November 2019. Figure 4 shows that the number of EU websites with *Legitimate interest or No Response* decreases from 213 to 134 websites over the three periods. Most of this drop is explained by an increasing number of websites using consent mechanisms in response to the GDPR (the number increases from 199 to 285). The Stops EU Tracking response stays mostly stable over time. As for the US websites, (Figure 5 illustrates that their responses are much more stable over time. To summarize, the EU websites seem to adjust their response over time, with the majority of them ending up implementing a consent mechanism in response to GDPR. Ultimately, response trends are quite different between the EU and US websites.

Fig. 4 *Sankey Diagram: Evolution of EU Website-Level Responses over Time*

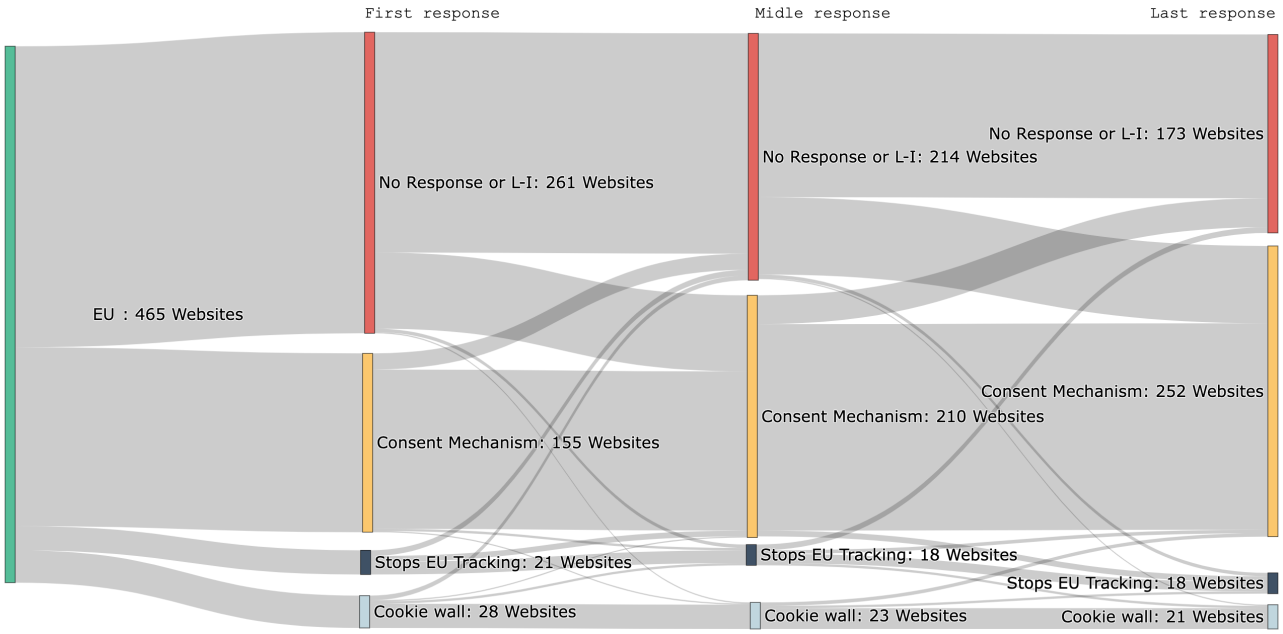
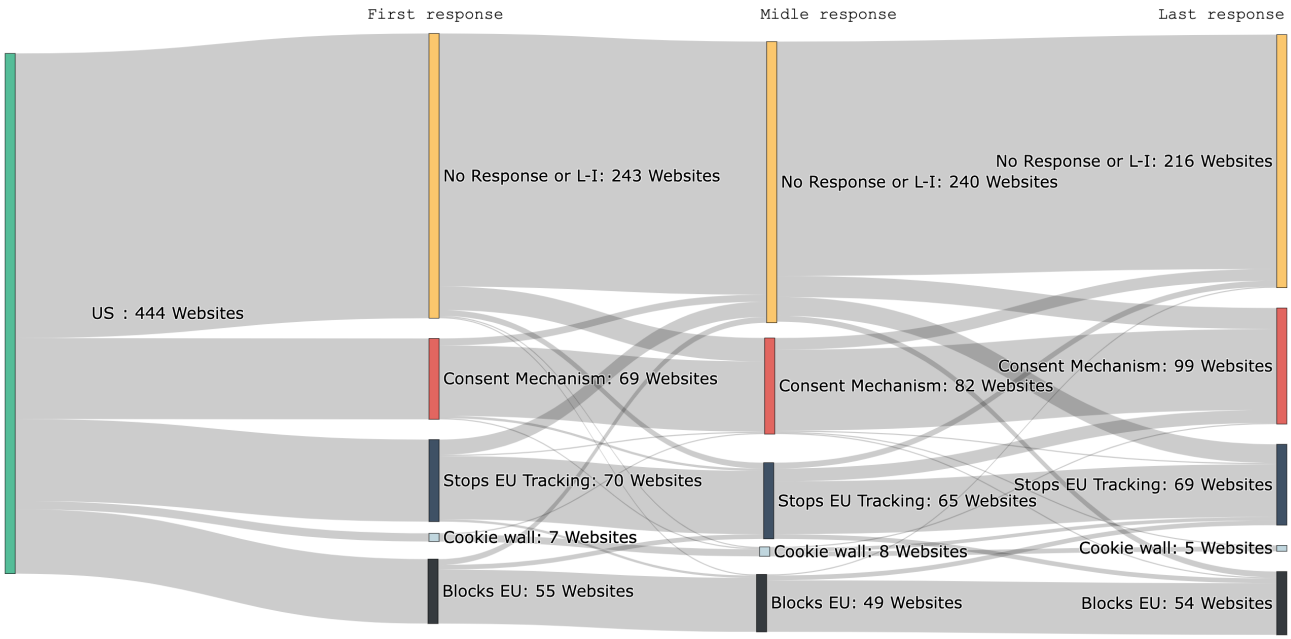


Fig. 5 *Sankey Diagram: Evolutions of US Website-Level Responses over Time*

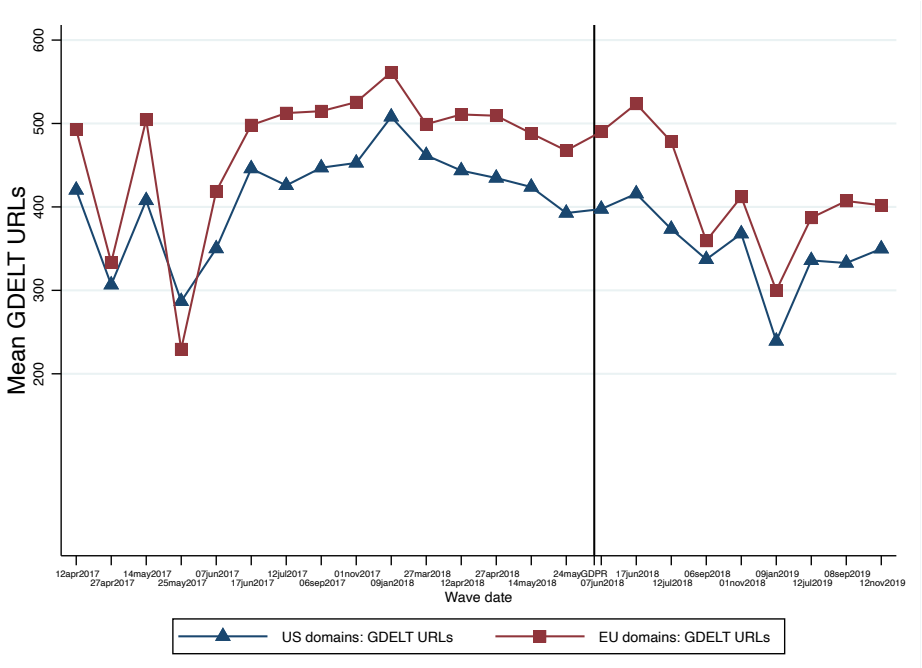


5.4 Changes in Downstream Variables

5.4.1 Content Quantity

We measure the impact of the GDPR on content quantity by capturing the number of new URLs of content posted by the websites in our sample (*GDELT URLs*). Figure 6 shows a similar initial decline in the number of new URLs published by both EU and US websites immediately after the enactment of the GDPR. Considering that the vast majority of US websites did not receive more than 2% of traffic from EU visitors, the generalized decline is likely either seasonal or due to factors (such as competition from streaming services) other than the GDPR. The number of new URLs starts recovering for both EU and US websites a few months following the enactment of the GDPR.

Fig. 6 *New URLs (GDELT URLs)*



Notes: The variable is the number of GDELT URLs and not the log

5.4.2 Content Quality

As proxies for content quality, we measure changes in websites' traffic (*Page Views Per User* and *Rank*) and social media engagement (the number of Facebook media "reactions" to new content published on the websites in our sample: *FB Average Reactions*).

Figure 7a shows the number of page views per user. While the trend seems stable for US websites, there appears to be a small downward trend for EU websites. A possible explanation for this is that the implementation of consent mechanisms, which we observe is the response adopted by the majority of websites in the EU over the long term, may have a negative effect on users' engagement. This could happen if, for example, consent pop-ups are considered obtrusive and time consuming, and some users turn away when they encounter them. In the case of the rank (Figure 7b) we observe that both EU and US websites experience a move towards lower positions (the result of less website traffic relative to other sites). In terms of reactions on Facebook, Figure 8 shows, initially, a stable trend, followed by alternating trends in later periods.

Fig. 7 *Websites Traffic*

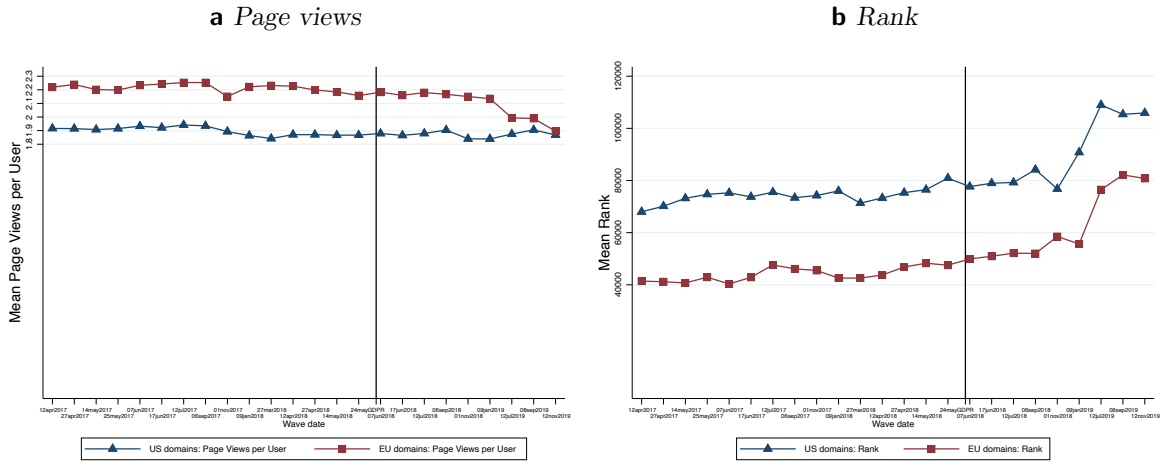
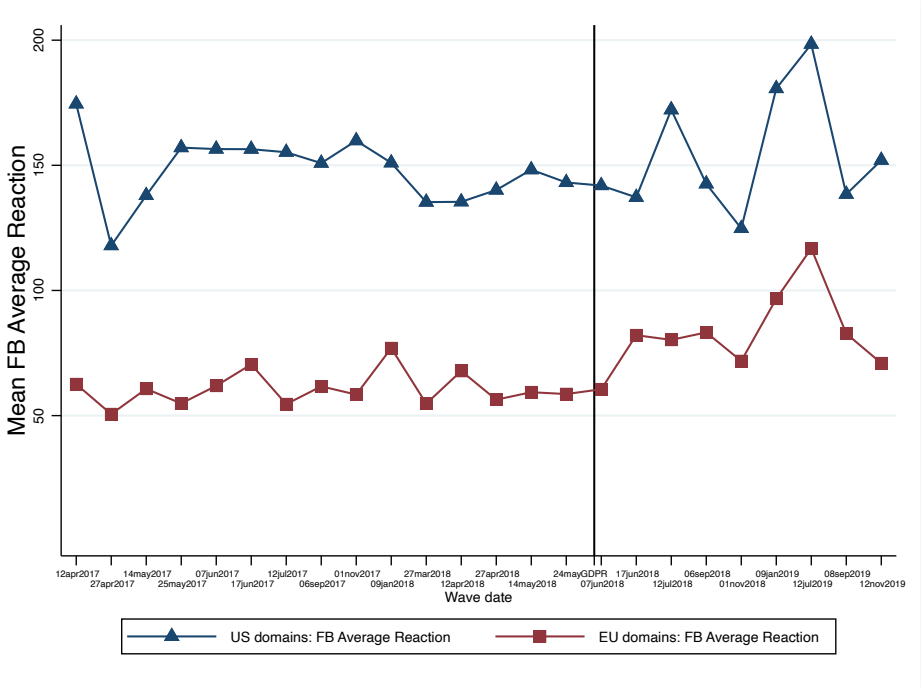


Fig. 8 *Facebook Reactions*



5.4.3 Website Survival

Only a very small minority of websites stopped producing content or shut down during the period of observation (around 1%). Based on GDELT data, we find that 4 websites in the EU sample and 6 websites in the US sample were no longer posting new content URLs as of Novemenber 2019.

6 Empirical Analysis

6.1 Identification Strategy

The descriptive evidence we have presented so far suggests not only significant changes in websites' handling of visitors' data following the GDPR, but nuanced and complex variations in websites' responses. Our objectives in the following sections are twofold: i) provide an estimate of the overall impact of the GDPR on content providers, by analyzing the impact on relevant downstream economic variables; and ii) provide evidence

on how the different websites responses (described above) may affect such variables of interest.

Our empirical approach relies on the fact that the GDPR can be thought of, to some extent, as an exogenous shock which, given its scope, should impact (more) certain websites than others. This makes it a good setting to implement a difference-in-differences approach, where we compare changes in outcome(s) for websites that have been (exogenously) impacted by the GDPR to the changes in outcome(s) for websites that should have not been impacted (or not impacted as much). Nevertheless, the general scope of the GDPR and the ubiquitous nature of the world wide web - where visitors from all over the world can browse and consume content from any website on the Internet - makes the context of our analysis challenging, for a number of reasons.

First, as outlined in the previous sections, we would expect the GDPR to affect, foremost, EU websites, which are subjected to the GDPR as EU entities (ecosystem effects; Section 3.1.1). Nevertheless, the GDPR applies to EU citizens as well, and, therefore, should apply to all websites (also not EU based) that have a considerable proportion of EU visitors. This makes it hard defining a clear "treatment" and a clear "control" group. Second, the GDPR enactment will produce the expected changes if websites do respond to it and implement the relevant changes. In fact, the generality of the GDPR, which leaves different options to the individual websites in terms of how to practically implement the regulation, implies that websites may interpret and respond in various ways; and each type of response may have potentially different implications, both in terms of outcome(s) as well as in terms of the strength of the response (website-level effects; Section 3.2). Both the decision to respond to the regulation, as well as the decision about how to respond are endogenous decisions of the individual websites, which can, therefore, be correlated to websites' observable and unobservable characteristics, adding an additional layer of complexity to our analysis.

Our empirical approach attempts to address the outlined challenges as follows. We start with a difference-in-differences (DID) approach aimed at estimating the overall impact of (the enactment) of the GDPR on the outcomes of interest. We begin by using a definition of treatment and control groups based on the geographical location of websites, where all the EU websites in our sample are considered as treated and all US websites in our sample are considered as controls. The effect estimated by this model is an intention to treat effect, since the estimation includes all the websites subjected to the initial treatment assignment, regardless of whether the websites did practically implement the GDPR or not. The estimate we obtain from this model gives us a measure of what we referred to as ecosystem effects: we measure the impact of the GDPR for EU websites, relative to US websites, regardless of whether they responded to the GDPR or not, and regardless the type of response potentially implemented. We then repeat the DID analysis by using a definition of treatment and control that considers both the geographical location of the websites and the geographical location of the visitors: we include among the treated all EU-based websites as well as US websites with a considerable proportion of EU visitors (details provided in the following section).

The intention to treat analysis does not take into consideration the fact that not all the websites respond to the enactment of the GDPR. In other words, the enactment of the GDPR can create not only ecosystem effects, but also websites-level effects, determined by whether and how websites respond. Given the challenges described above due to the endogeneity of websites' responses, we tackle this analysis using different strategies.

First, we use an instrumental variable (IV) approach aimed at estimating a local average treatment effect (LATE) - that is the effect of the GDPR for those websites that do respond to the regulation, regardless of the type of response. Next, we take into account that the predominant response over time for EU sites is the adoption of consent mechanisms and attempt to estimate the effect of that specific response, instead of any

response as we do in the LATE analysis. Attempting to estimate the consequences of a specific response is particularly problematic due to endogeneity concerns. To tackle this challenge we do an analysis based on the Look Ahead Propensity Score Matching methodology (Bapna *et al.*, 2018). Finally, we leverage the heterogeneity of the websites in our sample to explore if any of the effects we found are particularly prevalent for some groups of websites.

6.2 Naive difference-in-differences or Intention To treat

We use a difference-in-differences (DID) model to tease out potential changes in content quantity and user engagement after the GDPR, for EU websites, relative to US websites. Our framework controls for website specific features and time-specific characteristics. The specification of our regressions is as follows:

$$Y_{i,t} = \beta_0 + \beta_1 Post\ GDPR \times EU\ Websites_{i,t} + \omega_t + \mu_i + \epsilon_i \tag{1}$$

where, $Y_{i,t}$ represents our variable of interest for a website i at wave t; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects. $Post\ GDPR \times EU\ Websites_{i,t}$ is equal to 1 if the website i is a EU website and wave t was collected after the GDPR became effective, and 0 otherwise. Standard errors ϵ_i are clustered at the website level. The coefficient β_1 corresponds to the DID estimator of the effect of the implementation of the GDPR for websites based in the EU, compared to US websites.

The results are presented in Table 3. Column (1) presents the results for GDELT URLs that we use as proxy for content quantity. We use a logarithmic transformation of the dependant variable, *Log GDELT URLs*, to take into account the fact that our dependent variables is a count of new URLs. Columns (2) to (4) present the results for *Page Views Per User*, *Rank*, and *FB Average Reactions*, which we use as proxies for user engagement.

Table 3 is separated into three panels, each presenting the results for the analysis implemented using different definitions of control and treatment groups. The first panel shows the results for our basic specification, where the treatment group includes all EU based websites while the control group includes all US based websites. In the second panel, the treatment group includes EU based websites and US websites with a share of EU visitors greater than 10%. In the last panel, the treatment group includes EU based websites and US websites with a share of EU visitors greater than 5%. The implementation of these different definitions of treatment and control aims at mitigating concerns due to the fact that the broad application of the GDPR can lead some US websites to respond to it, where the degree of response should be correlated to the amount of EU visitors received by those websites.

The results obtained are consistent across the panels. We find a negative and statistically significant effect for *Page Views Per User* (Column 2), suggesting that, after the enactment of the GDPR, EU websites experience a decrease in the number of pages browsed by their visitors. One possible interpretation is that the reduction in the number of pages visited on a given website may be a signal of reduction in the quality of the content offering. If the quality of the content is reduced, users may decide to spend less time on the website and divert their attention to other websites. Another plausible explanation is that when users visit EU websites, they now encounter consent notices or requests, and in some cases even cookie walls. This may lead viewers to leave the page instead of expressing their consent choices. While the results suggest a negative impact on page views, this does not seem to translate to a deterioration of the ranking of websites, as we do not find any statistically significant effect of the GDPR on rank (Column 3). We do not find a statistically significant effect for GDELT URLs (1), nor Facebook reactions (4) (one of our proxies for user engagement).

In summary, the results suggest that the enactment of the GDPR may have, to some extent, negatively affected page views per user for EU websites, relative to US

websites, but has not relatively affected the amount of content that EU websites are able to publish, or the degree of average social media engagement and interaction with such content. Our expectation of these variables is that they are strongly correlated with content quality, considering that the goal of these websites is to attract and retain viewers.

Table 3: Diff-in-diff Estimations

	Content Quantity	User Engagement		
	(1)	(2)	(3)	(4)
	Log GDELT URLs	Page Views Per User	Rank	FB Average Reaction
Intention to Treat 1: Treatment group base on EU Websites				
EU Websites × Post GDPR	0.006	-0.092***	2539.370	11.980
	(0.041)	(0.033)	(3322.068)	(15.023)
Constant	5.015***	2.049***	64575.950***	108.380***
	(0.007)	(0.006)	(637.224)	(2.670)
Website fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster
Obs.	17577	21797	21797	17577
Intention to Treat 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors				
EU Websites and > 10% EU Visitors × Post GDPR	0.004	-0.119***	2529.713	13.988
	(0.041)	(0.032)	(3326.240)	(14.867)
Constant	5.015***	2.054***	64588.247***	108.081***
	(0.007)	(0.006)	(624.290)	(2.581)
Website fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster
Obs.	17577	21797	21797	17577
Intention to Treat 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors				
EU Websites and > 5% EU Visitors × Post GDPR	0.004	-0.114***	3172.219	15.511
	(0.041)	(0.032)	(3323.893)	(14.912)
Constant	5.015***	2.053***	64463.728***	107.797***
	(0.007)	(0.006)	(627.967)	(2.607)
Website fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster
Obs.	17577	21797	21797	17577

Notes: Estimates from the DID estimation for all websites in our sample. Column (1) presents estimation for the *Log Number of GDELT URL* dependent variable. Column (2) reports estimation with *Page Views Per User* as dependent variable. Column (3) presents *Rank* as dependent variable. Column (4) reports estimation with *FB average Reaction* as dependent variable. All estimations include time and website fixed effect.

Standard errors in parentheses and clustered at the website level.

Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.3 LATE

The analysis presented above gives us an estimation of the overall ecosystem effect of the GDPR. In this section, we use an instrumental variable (IV) approach to estimate the effect of the GDPR for the websites that do respond to the regulation, regardless of the type of response. As a consequence, the effect estimated represents a local average treatment effect (LATE) or average effects for those websites which decide to respond to the GDPR. This effect corresponds to a combination of website-level response and ecosystem effects. While the decision to respond to the GDPR is an endogenous decision of the websites, we exploit the fact that the (exogenous) enactment of the GDPR can be used as an instrument for the decision of a website to respond to the regulation, an approach first discussed by Angrist and Imbens (1995). To implement this approach, we need to identify if a website is responding to the GDPR or not. We use a conservative approach and assume that a website is responding to the GDPR if it implements a response which is clearly visible to the visitor. This includes websites that implement a consent mechanism; that stops tracking users; that implement a cookie wall or a cookie banner; and websites that block EU users (applies to US websites only).

In our first stage specification, we therefore estimate the probability of a website to respond to the GDPR as function of the enactment of the the GDPR.

$$GDPR\ Response_{i,t} = \alpha_0 + \alpha_1 Post\ GDPR \times EU\ Websites_{i,t} + \omega_t + \mu_i + \zeta_i \qquad (2)$$

Where, $GDPR\ Response_{i,t}$ is equal to 1 if the website responded to the GDPR (based on the definition of response provided above) and 0 otherwise; $Post\ GDPR \times EU\ Websites_{i,t}$ is our instrument; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects and ζ_i is the error. In the second stage, we use the residuals obtained from the first stage to estimate the impact of responding to the GDPR, on our outcomes of interest: *Log GDELT URLs*, *Page Views Per User*, *Rank*, and *FB Average Reactions*.

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{GDPR\ Response}_{i,t} + \omega_t + \mu_i + \epsilon_i \quad (3)$$

where $Y_{i,t}$ represents our variable of interest for a website i at wave t; $\widehat{GDPR\ Response}_{i,t}$ is the estimation of response from the first stage; ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects.

Table 4 presents the results of our IV approach. Similarly to the DID analysis, we use three different definitions of treatment and control (see Section 6.2). Results are consistent across the different panels: for all our outcomes of interest, we do not find a statistically significant effect of the GDPR for websites that do choose to respond to the regulation.

Table 4: LATE Estimations

	Log GDELT URLs		Page Views Per User		Rank		FB Average Reaction	
	(1) First Stage GDPR Response	(2) 2SLS Log GDELT URLs	(3) First Stage GDPR Response	(4) 2SLS Page Views Per User	(5) First Stage GDPR Response	(6) 2SLS Rank	(7) First Stage GDPR Response	(8) 2SLS FB Average Reaction
LATE 1: Treatment group base on EU Websites								
EU websites × Post GDPR	-0.005		0.019		0.019		-0.005	
	(0.022)		(0.019)		(0.019)		(0.022)	
GDPR Response		-1.159		-4.856		1.34e+05		-2518.916
		(9.886)		(5.232)		(2.19e+05)		(11827.717)
Constant	0.001		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.681		0.681		0.681		0.681
Share of response inside Treatment		0.816		0.816		0.816		0.816
Underidentification (LM)		0.048		1.002		1.002		0.048
P-value (LM-Stat)		0.826		0.317		0.317		0.826
Weak identification		0.048		1.001		1.001		0.048
P-value (J-Stat)		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577
LATE 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors								
EU Websites and > 10% EU Visitors × Post GDPR	-0.002		0.020		0.020		-0.002	
	(0.022)		(0.019)		(0.019)		(0.022)	
GDPR Response		-1.713		-5.915		1.26e+05		-6482.808
		(24.944)		(5.832)		(2.01e+05)		(64855.586)
Constant	0.001		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.682		0.682		0.682		0.682
Share of response inside Treatment		0.818		0.818		0.818		0.818
Underidentification (LM)		0.010		1.148		1.148		0.010
P-value (LM-Stat)		0.920		0.284		0.284		0.920
Weak identification		0.010		1.147		1.147		0.010
P-value (J-Stat)		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577
LATE 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors								
EU Websites and > 5% EU Visitors × Post GDPR	-0.004		0.019		0.019		-0.004	
	(0.022)		(0.019)		(0.019)		(0.022)	
GDPR Response		-0.947		-6.085		1.69e+05		-3621.372
		(10.477)		(6.444)		(2.44e+05)		(18483.406)
Constant	0.001		0.000		0.000		0.001	
	(0.006)		(0.005)		(0.005)		(0.006)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.682		0.682		0.682		0.682
Share of response inside Treatment		0.817		0.817		0.817		0.817
Underidentification (LM)		0.040		0.990		0.990		0.040
P-value (LM-Stat)		0.842		0.320		0.320		0.842
Weak identification		0.039		0.989		0.989		0.039
P-value (J-Stat)		0.000		0.000		0.000		0.000
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577

Estimates from the LATE estimation.
Standard errors in parentheses and clustered at the website level.
Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

Nevertheless, we suspect that the extent to which a website is going to respond and the effect of such response may vary with the website's share of EU visitors. Stated differently, websites with a larger share of EU visitors should have more incentive to respond to the regulation and might be more impacted since his response concerns a larger share of consumer. To explore this idea, we repeat our IV analysis by interacting the website response with the website's share of EU visitors. Results are reported in Table 5. The findings, still consistent across the different definitions of treatment and control, suggest that websites that do respond to the GDPR experience a statistically significant decrease in *Page Views Per User*. The decrease tends to be larger (in magnitude) the larger the website's share of EU visitors.

Table 5: LATE Estimations Accounting for Share of EU Visitors

	Log GDELT URLs		Page Views Per User		Rank		FB Average Reaction	
	(1) First Stage GDPR Response × Share EU visitors	(2) 2SLS Log GDELT URLs	(3) First Stage GDPR Response × Share EU visitors	(4) 2SLS Page Views Per User	(5) First Stage GDPR Response × Share EU visitors	(6) 2SLS Rank	(7) First Stage GDPR Response × Share EU visitors	(8) 2SLS FB Average Reaction
LATE 1: Treatment group base on EU Websites								
EU websites × Post GDPR × Share EU visitors	-0.144***		-0.103***		-0.103***		-0.144***	
	(0.034)		(0.027)		(0.027)		(0.034)	
'Post GDPR × Share EU Visitors	0.709***		0.694***		0.694***		0.709***	
	(0.018)		(0.015)		(0.015)		(0.018)	
EU websites × Post GDPR	0.052***		0.037***		0.037***		0.052***	
	(0.015)		(0.012)		(0.012)		(0.015)	
GDPR Response		0.811		0.197		-5.13e+04		85.970
		(0.654)		(1.132)		(59814.548)		(325.756)
GDPR Response × Share EU Visitors		0.036		-0.210***		10023.074		12.971
		(0.089)		(0.067)		(7244.311)		(23.957)
Constant	0.001		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.681		0.681		0.681		0.681
Share of response inside Treatment		0.816		0.816		0.816		0.816
Underidentification (LM)		11.583		9.825		9.825		11.583
P-value (LM-Stat)		0.003		0.007		0.007		0.003
Weak identification		1.755		2.634		2.634		1.755
J-Stat		0.438		0.177		3.285		0.730
P-value (J-Stat)		0.508		0.674		0.070		0.393
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577
LATE 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors								
EU Websites and > 10% EU Visitors × Post GDPR × Share EU Visitor	-0.164***		-0.116***		-0.116***		-0.164***	
	(0.037)		(0.031)		(0.031)		(0.037)	
'Post GDPR × Share EU Visitors	0.707***		0.693***		0.693***		0.707***	
	(0.017)		(0.015)		(0.015)		(0.017)	
EU Websites and > 10% EU Visitors × Post GDPR	0.071***		0.050***		0.050***		0.071***	
	(0.019)		(0.016)		(0.016)		(0.019)	
GDPR Response		0.558		-1.011		-6.77e+04		265.913
		(0.633)		(0.651)		(44220.347)		(379.069)
GDPR Response × Share EU Visitors		0.030		-0.190***		10306.816		17.196
		(0.084)		(0.073)		(7367.417)		(27.219)
Constant	0.001		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.682		0.682		0.682		0.682
Share of response inside Treatment		0.818		0.818		0.818		0.818
Underidentification (LM)		14.723		14.543		14.543		14.723
P-value (LM-Stat)		0.001		0.001		0.001		0.001
Weak identification		3.325		4.551		4.551		3.325
J-Stat		1.230		0.241		4.883		1.243
P-value (J-Stat)		0.267		0.623		0.027		0.265
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577
LATE 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors								
EU Websites and > 5% EU Visitors × Post GDPR × Share EU Visitor	-0.157***		-0.112***		-0.112***		-0.157***	
	(0.037)		(0.030)		(0.030)		(0.037)	
Post GDPR × Share EU Visitors	0.708***		0.694***		0.694***		0.708***	
	(0.018)		(0.015)		(0.015)		(0.018)	
EU Websites and > 5% EU Visitors × Post GDPR	0.063***		0.045***		0.045***		0.063***	
	(0.018)		(0.015)		(0.015)		(0.018)	
GDPR Response		0.706		-0.829		-3.13e+04		261.759
		(0.640)		(0.731)		(58780.399)		(360.743)
GDPR Response × Share EU Visitors		0.033		-0.193***		9677.160		17.374
		(0.087)		(0.070)		(7129.807)		(26.850)
Constant	0.001		0.000		0.000		0.001	
	(0.004)		(0.004)		(0.004)		(0.004)	
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Share of response inside Control		0.682		0.682		0.682		0.682
Share of response inside Treatment		0.817		0.817		0.817		0.817
Underidentification (LM)		12.701		11.524		11.524		12.701
P-value (LM-Stat)		0.002		0.003		0.003		0.002
Weak identification		2.226		3.260		3.260		2.226
J-Stat		0.822		0.115		2.235		1.604
P-value (J-Stat)		0.364		0.734		0.135		0.205
Obs.	17,588	17,577	21,797	21,797	21,797	21,797	17,588	17,577

Estimates from the LATE estimation.
Standard errors in parentheses and clustered at the website level.
Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.4 Lookahead Matching

In both the difference-in-difference and LATE analysis, the estimates we obtain capture both ecosystem and website-level response effects. In this section we attempt to isolate the effect of website-level responses. We focus on the implementations of consent mechanisms, as this is the most prevalent response in the long-term as shown in figure 4. This analysis is challenging because the response websites implement is very

likely correlated with its characteristics, many of which may not be observable. To deal with this challenge we implement an analysis following the logic of the Look Ahead Propensity Score Matching methodology (Bapna *et al.*, 2018). The endogeneity problem prevents us from directly comparing websites that adopt vs. websites that don't adopt a particular response, as the differences we observe may be driven by the factors that led adopters to implement the response rather than by the response itself. The look ahead matching methodology is based on comparing websites that have adopted a response with websites that have not adopted a response but will adopt it some time in the future. In this way we isolate ourselves from the endogeneity problem, as we are basing our analysis only on websites that will end up adopting a response and exploit the temporal variation in adoption to identify the impact of the response on our variables of interest.

Specifically, for this analysis we focus only on EU websites during the post-GDPR period that adopt a consent mechanism at some point in time. Of the 466 EU websites in our sample, 209 adopt a consent mechanism for at least one of the waves. Considering only these websites, we estimate a linear regression of the form:

$$Y_{i,t} = \beta_0 + \beta_1 \times CM_{i,t} + \beta_2 \times CM_{i,t} \times Share\ EU_i + \omega_t + \mu_i + \epsilon_i \tag{4}$$

In this equation, $Y_{i,t}$ corresponds to the outcomes we study for website i at time t, CM is equal to 1 if website i is showing a consent mechanism to EU visitors at time tT, $Share\ EU_i$ corresponds to the fraction of visitors website i gets from the EU (which should correspond to the share of users that observe the consent mechanism), ω_t is a vector of time fixed effects, and μ_i is a vector of website fixed effects. Standard errors are clustered at the website level. In this estimation β_2 corresponds to the effect of consent mechanisms on our outcome variables of interest.

Table 6: Look Ahead Matching

	Content Quantity	User Engagement		
	(1)	(2)	(3)	(4)
	Log GDELT URLs	Page Views Per User	Rank	FB Average Reaction
Consent Mechanism	0.0601	0.194**	-6801	210.2
	(0.199)	(0.0806)	(10697)	(165.5)
Consent Mechanism × Share EU	-0.121	-0.281***	-3422	-227.9
	(0.225)	(0.107)	(17816)	(184.9)
Constant	4.945***	2.031***	62,366***	33.87
	(0.0492)	(0.0226)	(5991)	(27.00)
Website fixed effect	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster
Observations	1434	1878	1878	1434
R-squared	0.905	0.722	0.713	0.529

Notes: Estimates from the Look Ahead Matching estimation. Includes only EU websites, after the GDPR, and that were using a consent mechanism in at least one of the waves. Column (1) presents estimation for the *Log Number of GDELT URL* dependent variable. Column (2) reports estimation with *Page Views Per User* as dependent variable. Column (3) presents *Rank* as dependent variable. Column (4) reports estimation with *FB average Reaction* as dependent variable. All estimations include time and website fixed effect.

Standard errors in parentheses and clustered at the website level.

Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

The results presented in table 6 show that the only statistically significant effect of the presence of a consent mechanism is over page views per user. The coefficient estimates imply that a EU website that received all of its visitors from the EU and implements a consent mechanism would see a decrease of 0.09 page views per user. This is a smaller effect than what we found in our LATE estimations that implied a reduction in the order of 0.2 page views per user. This suggests that the effect of website-level responses don't explain all of the impact of the GDPR on page views per user that we identified through the LATE analysis. This suggests the presence of both ecosystem and website-level effects after the GDPR. In this analysis we employ the difference-in-difference specifications.

6.5 Heterogeneous effect

In order to measure the effect of the GDPR on websites' ability to provide high quality content, we focused on websites in the news and media category. We only found an effect on the page views per user, but none on content quantity. However, within the news websites there is a wide range of website heterogeneity. We use the heterogeneity of websites to understand which websites are most impacted and why. In order to explore this effect, we split the sample into groups of websites according to their pre-GDPR characteristic.

6.5.1 Advertising

We ran the same analysis for the sub-sample of websites that relied on advertising to monetize their content before the GDPR. We separate our sample in two groups—"low" and "high" ads—respectively representing websites below and above the median advertising length before the GDPR. Table 7 shows statistically significant reductions in the number of page views per user for EU websites that rely significantly on advertising before the GDPR no matters the treatment group. This suggests a decrease in traffic also for this type of EU website, relative to US websites, after the implementation of the GDPR. However, we find no statistically significant change in the quantity of new content published or in Facebook reactions.

Table 7: Diff-in-diff regressions: For content variables dependent by Low and High advertising subsample

	Log GDELT URLs		Page Views Per User		Rank		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Low Ads	High Ads	Low Ads	High Ads	Low Ads	High Ads	Low Ads	High Ads
Intention to Treat 1: Treatment group base on EU Websites								
Treated (EU Web) × Post GDPR	0.060	-0.015	-0.062	-0.135***	-52.711	2847.824	3.297	11.129
	(0.067)	(0.055)	(0.057)	(0.036)	(5657.050)	(3787.949)	(23.327)	(18.528)
Constant	4.662***	5.288***	2.064***	2.036***	68011.163***	60787.870***	128.258***	95.349***
	(0.014)	(0.008)	(0.012)	(0.006)	(1201.870)	(624.157)	(4.729)	(2.811)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	7819	9196
Intention to Treat 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors								
EU Websites and > 10% EU Visitors × Post GDPR	0.052	-0.015	-0.118**	-0.136***	54.909	2972.581	10.198	10.096
	(0.066)	(0.055)	(0.058)	(0.036)	(5636.333)	(3791.712)	(23.198)	(18.451)
Constant	4.664***	5.288***	2.074***	2.036***	67988.702***	60769.582***	126.940***	95.514***
	(0.013)	(0.008)	(0.012)	(0.006)	(1156.048)	(621.883)	(4.519)	(2.783)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	7819	9196
Intention to Treat 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors								
EU Websites and > 5% EU Visitors × Post GDPR	0.053	-0.015	-0.107*	-0.135***	1484.354	2847.824	12.512	11.129
	(0.066)	(0.055)	(0.058)	(0.036)	(5618.205)	(3787.949)	(23.259)	(18.528)
Constant	4.664***	5.288***	2.072***	2.036***	67692.996***	60787.870***	126.468***	95.349***
	(0.013)	(0.008)	(0.012)	(0.006)	(1161.858)	(624.157)	(4.569)	(2.811)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	7819	9196	10614	10481	10614	10481	7819	9196

Notes: Estimates from the DID estimation for website with high and low level advertising before the GDPR. Column (1) presents estimation for the *Log GDELT URLs* dependent variable on the low Ads subsample. Column (2) reports estimation with *Log GDELT URLs* as dependent variable for the High ads subsample. Column (3) and (4) present respectively *Page Views Per User* as dependent variable for low and high Ads subsample. Column (5) and (6) reports estimation with *FB average Reaction* as dependent variable. All estimations include waves and website fixed effect.

Standard errors in parentheses and clustered at the website level.

Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.5.2 Ranking

Given the heterogeneity of websites in terms of size and compliance abilities, we investigate how the GDPR has affected the most prominent websites according to Alexa ranking. We split our sample into two groups: the first group consists of websites ranking in the top 50% of sites in their respective region (EU/US) with respect to sites in

our sample; The second group includes the remaining websites. Table 8 presents the DID estimation of the effect the GDPR on content for the top and bottom EU websites. Columns (1), (3), (5) and (7) report the estimation for the sub-sample of the top websites. The results reported in Columns (3) and (4) suggest that the implementation of the GDPR has a negative and significant effect on page views for top and bottom EU websites. However, there does not seem to be a significant change in number of GDELT URLs (columns 1 and 2), or Facebook reactions (columns 7 and 8).

Table 8: Diff-in-diff regressions: based on the subsample of the top/bottom websites in EU and in US

	Log GDELT URLs		Page Views Per User		Rank		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Top	Bottom	Top	Bottom	Top	Bottom	Top	Bottom
Intention to Treat 1: Treatment group base on EU Websites								
Treated (EU Web) × Post GDPR	0.076	-0.062	-0.107**	-0.092***	1043.841	4442.115	17.379	0.502
	(0.059)	(0.059)	(0.054)	(0.035)	(3018.552)	(6009.096)	(28.517)	(7.085)
Constant	5.323***	4.634***	2.151***	1.949***	13239.239***	1.16e+05***	170.764***	40.099***
	(0.011)	(0.010)	(0.010)	(0.007)	(572.214)	(1129.430)	(5.223)	(1.176)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	9096	7919	10619	10476	10619	10476	9096	7919
Intention to Treat 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors								
EU Websites and > 10% EU visitors × Post GDPR	0.069	-0.059	-0.162***	-0.092***	-260.787	5291.371	21.319	0.805
	(0.059)	(0.059)	(0.053)	(0.035)	(3024.609)	(6011.410)	(28.096)	(7.042)
Constant	5.324***	4.633***	2.161***	1.949***	13485.054***	1.16e+05***	170.162***	40.051***
	(0.010)	(0.010)	(0.010)	(0.006)	(555.988)	(1115.519)	(4.989)	(1.152)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	9096	7919	10619	10476	10619	10476	9096	7919
Intention to Treat 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors								
EU Websites and > 5% EU visitors × Post GDPR	0.069	-0.059	-0.151***	-0.092***	1334.354	5291.371	24.126	0.805
	(0.059)	(0.059)	(0.054)	(0.035)	(3020.706)	(6011.410)	(28.270)	(7.042)
Constant	5.324***	4.633***	2.159***	1.949***	13188.566***	1.16e+05***	169.608***	40.051***
	(0.011)	(0.010)	(0.010)	(0.006)	(562.667)	(1115.519)	(5.085)	(1.152)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	9096	7919	10619	10476	10619	10476	9096	7919

Notes: Estimates from the DID estimation for website with top and small rank before the GDPR. Column (1) presents estimation for the *Log GDELT URLs* dependent variable on the top rank subsample. Column (2) reports estimation with *Log GDELT URLs* as dependent variable for the small rank subsample. Column (3) and (4) present respectively *Page Views Per User* as dependent variable for top and small rank subsample. Column (5) and (6) reports estimation with *FB average Reaction* as dependent variable. All estimations include waves and website fixed effect.

Standard errors in parentheses and clustered at the website level.

Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

6.5.3 Short vs Long run

The lack of consensus on what it means to be compliant with the GDPR may delay the effect (if any) of the regulation. As we show in Section 5, on the one hand, websites reacted rapidly to the enactment of the regulation by reducing the magnitude of visitor

tracking - however, such reduction did not last over time. On the other hand, responses for a majority of EU websites evolved over time. In this section we compare short and long-run effects. We split our sample after the GDPR into two groups. The short-run subsample presents the effect of the GDPR until the beginning of January 2019. Table 9 column (1), column(3), column (5) and column (7) present the estimation on the short-run group. The long-run subsample excludes from the analysis the time period just after the GDPR, meaning the period from June 2018 to January 2019. In other words, it includes all the pre-GDPR waves (from April 2017 to May 25, 2018) and the latest post-GDPR waves (from July 2019 to November 2019). Table 9 column (4) shows that the decrease in page views per user for EU websites compared to US websites seems to only be a long-run effect.

Table 9: Diff-in-diff regressions: For content variables dependent by short and long-run subsample

	Log GDELT URLs		Page Views Per User		Rank		FB Average Reaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Short Run	Long Run	Short Run	Long Run	Short Run	Long Run	Short Run	Long Run
Intention to Treat 1: Treatment group base on EU Websites								
Treated (EU Web) × Post GDPR	-0.006	0.019	-0.023	-0.229***	1922.883	3379.643	13.798	10.722
	(0.038)	(0.067)	(0.031)	(0.057)	(3287.165)	(5158.319)	(15.087)	(24.070)
Constant	5.037***	5.041***	2.051***	2.056***	60798.116***	64149.592***	106.497***	108.150***
	(0.005)	(0.005)	(0.004)	(0.005)	(480.746)	(439.013)	(2.058)	(1.819)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	15566	13222
Intention to Treat 2: Treatment group base on EU Websites + Websites with more than 10% of EU visitors								
EU Websites and > 10% EU visitors × Post GDPR	-0.010	0.024	-0.039	-0.279***	1043.402	5107.264	16.283	11.719
	(0.038)	(0.067)	(0.031)	(0.056)	(3311.108)	(5130.506)	(14.960)	(23.715)
Constant	5.037***	5.041***	2.053***	2.060***	60930.021***	64011.932***	106.210***	108.095***
	(0.005)	(0.005)	(0.004)	(0.005)	(473.834)	(427.228)	(1.992)	(1.751)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	15566	13222
Intention to Treat 3: Treatment group base on EU Websites + Websites with more than 5% of EU visitors								
EU Websites and > 5% EU visitors × Post GDPR	-0.010	0.024	-0.036	-0.270***	2234.922	4650.988	16.944	15.165
	(0.038)	(0.067)	(0.031)	(0.057)	(3302.734)	(5138.329)	(14.994)	(23.822)
Constant	5.037***	5.041***	2.053***	2.059***	60757.401***	64047.366***	106.107***	107.831***
	(0.005)	(0.005)	(0.004)	(0.005)	(475.752)	(430.709)	(2.010)	(1.773)
Website fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time fixed effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Std. err Websites level	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Obs.	15566	13222	19077	16344	19077	16344	15566	13222

Notes: Estimates from the DID estimation for website on the short and long-run. Column (1) presents estimation for the *Log GDELT URLs* dependent variable on the short-run. Column (2) reports estimation with *Log GDELT URLs* as dependent variable for the long-run. Column (3) and (4) present respectively *Page Views Per User* as dependent variable for short and long-run. Column (5) and (6) reports estimation with *FB average Reaction* as dependent variable. All estimations include waves and website fixed effect.

Standard errors in parentheses and clustered at the website level.

Significance levels: $*p < .10$, $**p < .05$, $***p < .01$.

7 Discussion and Limitations

Our analysis provides insights into the impact of the GDPR on websites' content quantity and users' engagement as a proxy for content quality. Overall, the GDPR has reduced the number of third-party cookies and tracking responses, suggesting decreased

tracking of users by websites. This decrease is more evident for EU visitors to US websites, indicating that US websites are taking a conservative approach when dealing with the requirements of the GDPR. Furthermore, the enactment of the GDPR may have to some extent negatively affected page views of EU websites, relative to US websites. However, and importantly, the enactment does not seem to have relatively affected the amount of content that EU websites were able to publish, the amount of traffic they received, or the degree of average social media engagement and interaction with such content.

Before concluding, we feel it is important to highlight some limitations of our analysis. While we are using multiple measures to capture content quantity and quality, they are only proxies that may not fully capture the potential effect of the GDPR. Additionally, while we have classified cookies and HTTP requests to identify tracking and advertising related activity, and devised a way to detect the presence of consent mechanisms, our technical variables are only capturing a part of the technical changes that are possible.

Finally, despite being over two years into the GDPR, it may still be too early to detect changes in the content produced by publishers. Firms, weighting the cost of compliance against potential fines that may result from enforcement actions, may be inclined to wait until EU authorities provide further clarification on the requirements for compliance. Others still may be justifying data collection and processing under the 'legitimate interest' clause of Article 6. Indeed, a December 2019 report by the Dutch Data Protection Authority found that many popular websites were still placing tracking cookies on the browsers of EU visitors (Authority, 2019). If a significant number of websites are currently not fully compliant with the GDPR requirements, this would make the impact of the regulation on publishers' content weaker and thus more difficult to detect. It's possible that future clarifications or enforcement actions by the EU will trigger smaller scale market shocks as publishers are steered towards

compliance in areas such as consent.

8 Conclusion

While previous work has focused on measuring the effects of the GDPR on advertising technologies (such as cookies), the present study attempts to assess the impact of the GDPR on ad-supported content publishers by tracking the potential downstream economic effects of the regulation. We captured a number of metrics related to tracking, traffic, and content variables over several months, both leading up to and immediately following the enforcement of the GDPR.

We examined these variables using descriptive statistics and DID estimations. The DID analysis examined the changes in our variables for US and EU websites viewed from US and EU visitor addresses. For websites viewed from the EU relative to websites viewed from the US, our results indicate a reduction in the variables often associated with tracking; we also observed some evidence of a negative impact of the regulation on the traffic of EU websites. However, we did not find significant evidence of a negative effect of the regulation on the amount of content that EU websites publish, or the degree of average social media engagement and interaction with such content. The robustness of this result was confirmed by using different methodologies to account for endogeneity concerns. Moreover, we explicitly attempted to identify and distinguish between ecosystem effects (that affect all EU websites regardless of their response) and website-level effects. Our results suggest the existence of both types of effects.

References

Acquisti, A., Taylor, C. and Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature*. 54(2), 442–92.

Adjerid, I., Acquisti, A., Telang, R., Padman, R. and Adler-Milstein, J. (2015). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science*. 62(4), 1042–1063.

Anderson, S. and Gabszewicz, J. (2006). The media and advertising: A tale of two-sided markets. In *Handbook of the Economics of Art and Culture*. (p. 568–614.). vol. 1. Elsevier B.V., Amsterdam.

Angrist, J. and Imbens, G. (1995). *Identification and estimation of local average treatment effects*.

Aridor, G., Che, Y.-K., Nelson, W. and Salz, T. (2020). The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR. *Available at SSRN*.

Athey, S., Calvano, E. and Gans, J. S. (2018). The impact of consumer multi-homing on advertising markets and media competition. *Management Science*. 64(4), 1574–1590.

Authority, D. D. P. (2019). *AP: veel websites vragen op onjuiste wijze toestemming voor plaatsen tracking cookies*. `https://autoriteitpersoonsgegevens.nl/nl/nieuws/ap-veel-websites-vragen-op-onjuiste-wijze-toestemming-voor-plaatsen-tracking-cookies`.

Autoriteit Persoonsgegevens (2019). *Websites moeten toegankelijk blijven bij weigeren tracking cookies*.

Bapna, R., Ramaprasad, J. and Umyarov, A. (2018). Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement? *MIS Q.* 42(3), 719–736. ISSN 0276-7783. doi:10.25300/MISQ/2018/13592. Retrievable at `https://doi.org/10.25300/MISQ/2018/13592`.

Beales, H. (2010). The value of behavioral targeting. *Network Advertising Initiative*. 1, 2010.

Brill, J. (2018). *Microsoft's commitment to GDPR, privacy and putting customers in control of their own data*.

Cagé, J., Hervé, N. and Viaud, M.-L. (2015). The production of information in an online world. *The Review of Economic Studies*.

Casadesus-Masanell, R. and Zhu, F. (2013). Business model innovation and competitive imitation: The case of sponsor-based business models. *Strategic Management Journal*. 34(4), 464–482.

Castro, D. (2010). *Stricter privacy regulations for online advertising will harm the free internet*. Technical report. Information Technology and Innovation Foundation.

Choi, J. P., Jeon, D.-S. and Kim, B.-C. (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*. 173, 113–124.

Commission nationale de l'informatique et des libertés (2020). *Cookies et autres traceurs : la CNIL publie des lignes directrices modificatives et sa recommandation*. Technical report. Retrievable at `https://www.cnil.fr/fr/cookies-et-autres-traceurs-la-cnil-publie-des-lignes-directrices-modificatives-et-sa-recommandation`.

Commission nationale de l'informatique et des libertés (2021). *Nouvelles règles pour les cookies et autres traceurs : bilan de l'accompagnement de la CNIL et actions à venir*. Technical report. Retrievable at `https://www.cnil.fr/fr/nouvelles-regles-cookies-et-autres-traceurs-bilan-accompagnement-cnil-actions-a-venir`.

Dabrowski, A., Merzdovnik, G., Ullrich, J., Sendera, G. and Weippl, E. (2019). Measuring cookies and web privacy in a post-gdpr world. In *International Conference on Passive and Active Network Measurement*. Springer, 258–270.

Davies, J. (2019). *After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue*. `https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/`. Accessed: 2019-05-08.

Degeling, M., Utz, C., Lentzsch, C., Hosseini, H., Schaub, F. and Holz, T. (2019). We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Network and Distributed Systems Security (NDSS) Symposium 2019*. 345–346.

Deloitte (2013). *Economic impact assessment of the proposed General Data Protection Regulation*. Technical report. Deloitte. Retrievable at `https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/about-deloitte/deloitte-uk-european-data-protection-tmt.pdf`.

Downes, L. (2018). GDPR and the End of the Internet's Grand Bargain. *Harvard Business Review*. Retrievable at `https://hbr.org/2018/04/gdpr-and-the-end-of-the-internets-grand-bargain`.

Englehardt, S. and Narayanan, A. (2016). Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 1388–1401.

Evans, D. S. (2009). The Online Advertising Industry: Economics, Evolution, and Privacy. *Journal of Economic Perspectives*. 23(3), 37–60.

Ferreira, L. N., Hong, I., Rutherford, A. and Cebrian, M. (2021). The small-world network of global protests. *Scientific Reports*. 11. ISSN 2045-2322. doi:10.1038/s41598-021-98628-y.

Gallea, Q. and Rohner, D. (2021). Globalization mitigates the risk of conflict caused by strategic territory. *Proceedings of the National Academy of Sciences*. 118. ISSN 0027-8424. doi:10.1073/pnas.2105624118.

Godinho de Matos, M. and Adjerid, I. (2021). Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. *Management Science*. ISSN 0025-1909. doi:10.1287/mnsc.2021.4054.

Goldberg, S., Johnson, G. and Shriver, S. (2021). Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes. *Available at SSRN 3421731*.

Goldfarb, A. (2004). Concentration in advertising-supported online markets: An empirical approach. *Econom. Innovation New Tech.* 13(6), 581–594.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization*. 44(2), 115–129.

Goldfarb, A. and Tucker, C. (2011). Online display advertising: Targeting and obtrusiveness. *Marketing Science*. 30(3), 389–404.

Goldfarb, A. and Tucker, C. (2012). Privacy and innovation. *Innovation policy and the economy*. 12(1), 65–90.

Goldfarb, A. and Tucker, C. E. (2010). Privacy regulation and online advertising. *Management science*. 57(1), 57–71.

Hils, M., Woods, D. W. and Böhme, R. (2020). Measuring the Emergence of Consent Management on the Web. In *Proceedings of the ACM Internet Measurement Conference*. 10. ACM. ISBN 9781450381383, 317–332. doi:10.1145/3419394.3423647.

IAB Europe (2021). *GDPR Guidance: Legitimate Interests Assessments (LIA) for Digital Advertising*.

IHS Technology (2015). *Paving the way: how online advertising enables the digital economy of the future*. Technical report.

International Association of Privacy Professionals (2020). *Behavioral Advertising*. `https://iapp.org/resources/article/behavioral-advertising-2/`.

Jia, J., Jin, G. Z. and Wagman, L. (2021). The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. *Marketing Science*, –. ISSN 0732-2399. doi:10.1287/mksc.2020.1271.

Johnson, G. and Shriver, S. (2019). Privacy & market concentration: Intended & unintended consequences of the GDPR. *Available at SSRN*.

Johnson, G. A., Lewis, R. A. and Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*. 54(6), 867–884.

Johnson, G. A., Shriver, S. K. and Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science*.

Kaye, K. (2021). *'We don't have visibility': Google's ad targeting limits expose publishers with reliance on open programmatic market and first-party data weakness*. Retrievable at `https://digiday.com/media/we-dont-have-visibility-googles-ad-targeting-limits-expose-publishers-with-reliance-on-open-programmatic-market-and-first-party-data-weakness/`.

Lambrecht, A., Goldfarb, A., Bonatti, A., Ghose, A., Goldstein, D. G., Lewis, R., Rao, A., Sahni, N. and Yao, S. (2014). How do firms make money selling digital goods online? *Marketing Letters*. 25(3), 331–341. doi:10.1007/s11002-014-9310-5.

Lancieri, F. (2021). Narrowing Data Protection's Enforcement Gap. *Available at SSRN 3806880*.

Lefouili, Y. and Toh, Y. L. (2018). Privacy Regulation and Quality Investment. *Working paper*.

Libert, T., Graves, L. and Nielsen, R. K. (2018). *Changes in Third-Party Content on European News Websites after GDPR*.

Marotta, V., Abhishek, V. and Acquisti, A. (2019). Online tracking and publishers' revenues: An empirical analysis. In *Workshop on the Economics of Information Security*.

Massé, E. (2021). *Three Years Under the EU GDPR: An Implementation Progress Report*. Retrievable at `https://www.accessnow.org/cms/assets/uploads/2021/05/Three-Years-Under-GDPR-report.pdf`.

Miller, A. R. and Tucker, C. (2009). Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records. *Management Science*. 55(7), 1077–9.

Monic, S. and Feng, Z. (2013). Ad Revenue and Content Commercialization: Evidence from Blogs. *Management Science*. 59(10), 2314–2331.

Nouwens, M., Liccardi, I., Veale, M., Karger, D. and Kagal, L. (2020). Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 4. ACM. ISBN 9781450367080, 1–13. doi:10.1145/3313831.3376321.

O'Reilly, L. (2020). *'Re-architecting the entire process': How Vice is preparing for life after the third-party cookie*. Retrievable at `https://digiday.com/media/re-architecting-the-entire-process-how-vice-is-preparing-for-life-after-the-third-party-cookie/`.

Peukert, C., Bechtold, S., Batikas, M. and Kretschmer, T. (2020). European Privacy Law and Global Markets for Data. *SSRN Electronic Journal*. ISSN 1556-5068. doi:10.2139/ssrn.3560392. Retrievable at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3560392`.

Phelan, D. (2018). *Apple Promotes Powerful Privacy Tools For iPhone, iPad, Mac Users In GDPR Response*.

Sakamoto, T. and Matsunaga, M. (2019). After GDPR, Still Tracking or Not? Understanding Opt-Out States for Online Behavioral Advertising. In *2019 IEEE Security and Privacy Workshops (SPW)*. 5. IEEE. ISBN 978-1-7281-3508-3, 92–99. doi:10.1109/SPW.2019.00027. Retrievable at `https://ieeexplore.ieee.org/document/8844599`.

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P.-A. and Santos, I. (2019). Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*. 340–351.

Sharma, P., Sun, Y. and Wagman, L. (2019). The Differential Effects of New Privacy Protections on Publisher and Advertiser Profitability. *SSRN Electronic Journal*. ISSN 1556-5068. doi:10.2139/ssrn.3503065. Retrievable at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3503065`.

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

Sørensen, J. and Kosta, S. (2019). Before and after gdpr: The changes in third party presence at public and private european websites. In *The World Wide Web Conference*. 1590–1600.

Tucker, C. (2012). The Economics of Advertising and Privacy. *International Journal of Industrial Organization*. 30(7).

UK Information Commissioner's Office (2019). *Update report into adtech and real time bidding*. `https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf`.

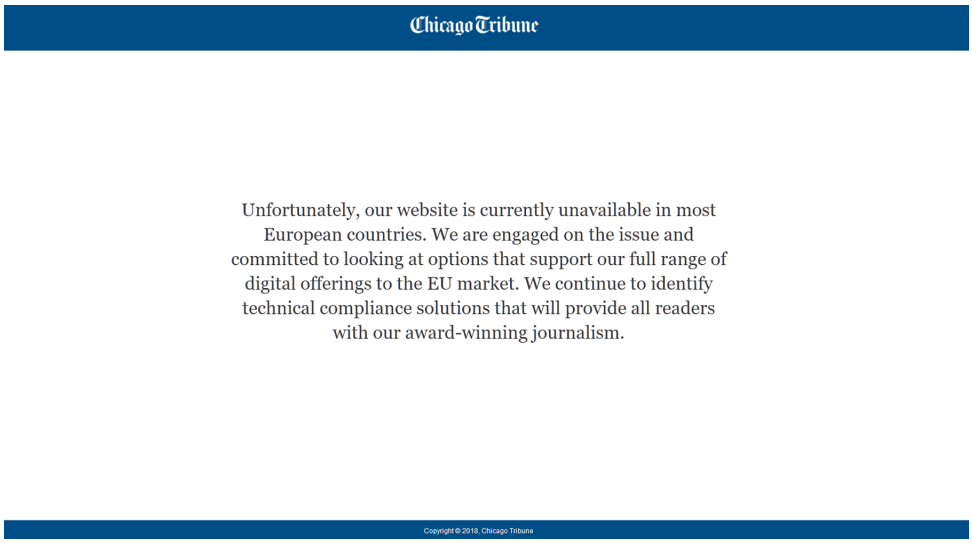
Urban, T., Tatang, D., Degeling, M., Holz, T. and Pohlmann, N. (2020). Measuring the Impact of the GDPR on Data Sharing in Ad Networks. *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*.

Utz, C., Degeling, M., Fahl, S., Schaub, F. and Holz, T. (2019). (Un) informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 973–990.

Zhuo, R., Huffaker, B., kc claffy and Greenstein, S. (2021). The impact of the General Data Protection Regulation on internet interconnection. *Telecommunications Policy*. 45. ISSN 03085961. doi:10.1016/j.telpol.2020.102083. Retrievable at `https://www.sciencedirect.com/science/article/abs/pii/S0308596120301737`.

9 Appendix

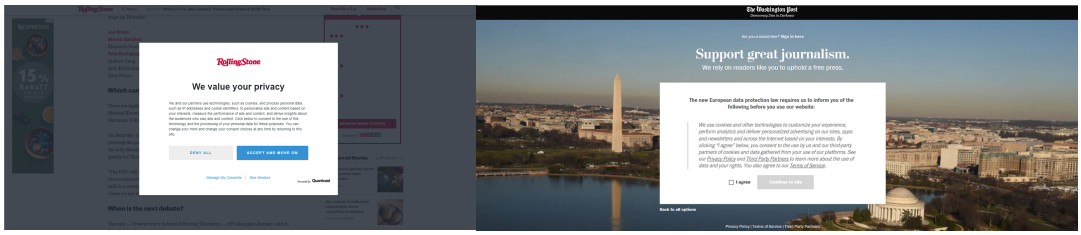
Fig. 9 *Example of Blocks EU Visitors*



Note: This figure presents an example of Blocks EU websites

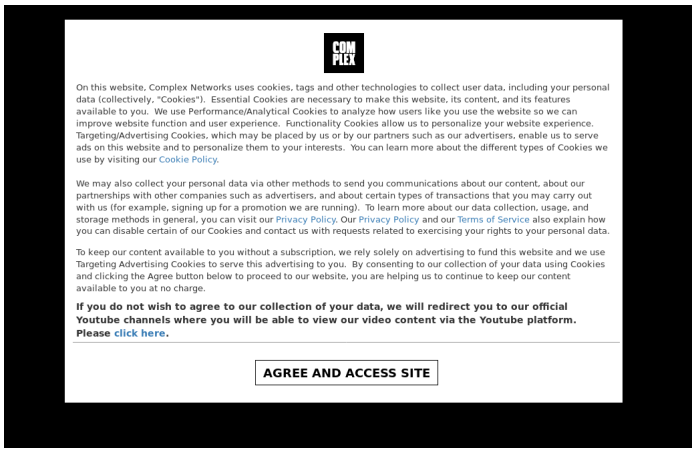
Fig. 10 *Examples of Types of Consent Mechanisms:*

a *With a direct Opt-out button* **b** *Without a direct Opt-out button*



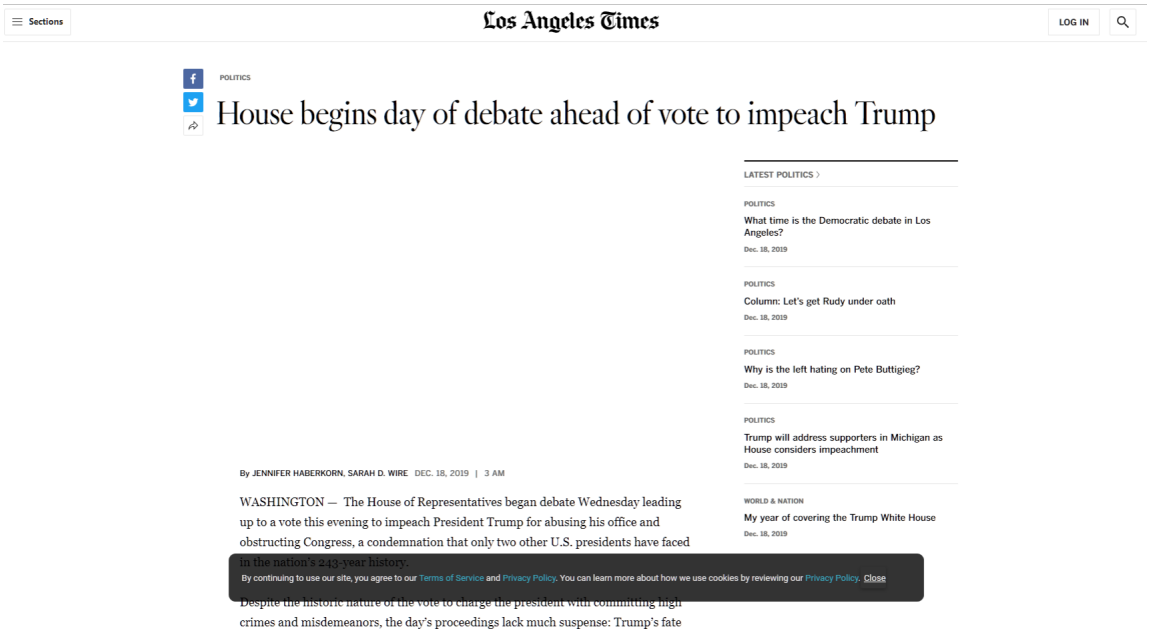
Note: This figure presents different kinds of consent mechanisms

Fig. 11 *Example of Cookie Wall*



Note: This figure presents a Cookie wall example

Fig. 12 *Cookies Notice*



Note: This figure presents different kinds of consent mechanisms