The Impact of Privacy Regulation on Web Traffic: Evidence From the GDPR [*]

Raffaele Congiu[†] Lorien Sabatino[†‡] Geza Sapi[§]

February 2022

Abstract

We use traffic data from around $5,000$ web domains in Europe and United States to investigate the effect of the European Union's General Data Protection Regulation (GDPR) on website visits and user behaviour. We document an overall traffic reduction of approximately 15% in the long-run and find a measurable reduction of engagement with websites. Traffic from direct visits, organic search, email marketing, social media links, display ads, and referrals dropped significantly, but paid search traffic – mainly Google search ads – was barely affected. We observe an inverted U-shaped relationship between website size and change in visits due to privacy regulation: the smallest and largest websites lost visitors, while medium ones were less affected. Our results are consistent with the view that users care about privacy and may defer visits in response to website data handling policies. Privacy regulation can impact market structure and may increase dependence on large advertising service providers. Enforcement matters as well: The effects were amplified considerably in the long-run, following the first significant fine issued eight months after the entry into force of the GDPR.

JEL Codes: D43, D8, L13, L15, L5

Keywords: Privacy regulation, GDPR, Website traffic, Marketing channels, Firm size

[*] We thank Similarweb (www.similarweb.com) for providing us the data used in this paper. We acknowledge financial support from the "Ministero dell'Istruzione, dell'Università e della Ricerca" Award TESUN - 83486178370409 finanziamento dipartimenti di eccellenza CAP. 1694 TIT. 232 ART. 6.

[†] Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

[‡] Corresponding author. E-mail: lorien.sabatino@polito.it.

[§] Düsseldorf Institute for Competition Economics (DICE), Heinrich Heine University of Düsseldorf. E-mail: sapi@dice.uni-duesseldorf.de. The views expressed in this article are solely those of the authors and may not, under any circumstances, be regarded as representing an official position of organizations the authors are affiliated with.

1 Introduction

The rise of the internet has enabled new products and services, revolutionizing the ways we work and interact. In the United States (US), as in other mature economies, the growth of the internet economy exceeds that of other sectors by orders of magnitude (Interactive Advertising Bureau, 2021). The shift of private and commercial activities to the internet goes hand in hand with increased attention to online security, privacy and the protection of data.[1] In recent years, policy makers around the world reacted to increased data protection concerns by putting in place a wave of regulations with the aim of protecting online and offline privacy, thereby limiting the ways companies collect and use data.[2] Typically, the goal of such privacy regulations is to empower consumers by giving them additional control over how firms gather and use their personal information.

One of the most far-reaching regulatory interventions to boost privacy is without doubt the General Data Protection Regulation (GDPR) of the European Union. The GDPR entered into force in May 2018 with the intention of protecting personal privacy across all domains, both online and offline. The regulation mandates data protection *"by design and by default"* and prescribes several principles and tools for managing data. It requires data controllers to inform individuals about the collection and use of data. Firms collecting personal data must obtain the users' informed opt-in consent to such practices. The regulation assigns liability to firms handling data and – for the first time – imposes significant fines on privacy breaches up to 4% of global turnover.

While the GDPR applies both online and offline to any entity handling personal information, no other economic domain is expected to be as heavily affected by it as the internet ecosystem. Today's internet relies heavily on the collection and use of visitor data. Web-

[1] In Europe, in 2011 around 40% of survey respondents were concerned about their behaviour being recorded through the internet when browsing, downloading files, and accessing content online (European Commission, 2011), in 2015 less than a quarter of respondents reported to trust online businesses to protect their personal data (European Commission, 2015). Growing online privacy concerns likely followed the general trend in internet use (Ourworldindata.org, 2021), and likely reach back to the early 2000s (Auxier et al., 2019).

[2] At the time of writing this article, the Australian Privacy Act of 1988 is under review (Coos, 2020), and legislation strengthening privacy is under way in Canada and the United States (Government of Canada, 2020; IAPP, 2021).

sites attract visitors by offering a broad spectrum of content ranging from news, music, and entertainment, to commerce and other services. More often than not, websites monetize at least in part by tracking their visitors and providing targeted advertisements to individuals or homogeneous groups of users. Digital stars such as Google and Facebook grew to titans of the internet by harvesting user data from millions of websites, connecting users' browsing paths, aggregating these signals into user profiles and offering advertising services to third parties to target consumers based on their specific interests and characteristics.[3]

Since its introduction, the GDPR has attracted considerable interest from researchers, policy-makers, and industry players across the globe.[4] So far, less attention has been devoted to understanding the consequences of the GDPR on website's ability to attract internet users and the altered ability and likelihood of those users to engage with website content. Our research aims to fill in this gap. We investigate the effect of the GDPR on online traffic and user behavior in the European Union (EU).

We use data on traffic broken down to detailed channels to around $5,000$ websites in Europe and the United States. We exploit the fact that the GDPR applies to European users, leaving the non-European audience unaffected. We apply a difference-in-differences (DiD) analysis that exploits the geographic origin of website traffic. In particular, our treatment assignment identifies the traffic originated from EU countries, leaving US traffic in the control group. This implies that multinational websites having visitors from both the EU and abroad are treated only for the portion of traffic coming from EU countries.

Overall, we find that the GDPR leads to a reduction in web traffic, and website visits decrease by approximately 15% in the long-run. This effect unfolds fully with a delay, several months after the hard date of the GDPR's entry into force, and following the issuance of the first large fine. We are able to break down the overall traffic reduction into various traffic acquisition channels. In the short-run, direct website traffic and visits triggered by email marketing messages reduced by 4.5% and 7% respectively, In the long-run – coinciding

[3] The tracker networks of Google and Facebook have been found embedded into respectively 80% and 37% of the most popular websites (Hu et al., 2020).

[4] The first research articles assessing the impact of the GDPR came from the computer science domain, and revolved around measuring websites' technical compliance with the regulation (Nouwens et al., 2020; Utz et al., 2019; Matte et al., 2020; Sørensen and Kosta, 2019).

with the first significant fine – traffic reduction amplified significantly. Email and display advertising traffic collapsed by 35% and 29%, respectively. Visits from referrals (such as links on 3rd party websites) and from social media reduce too, together with direct (−9.5%) and organic search visits (7%). Strikingly, against this backdrop we find that website traffic from paid search – mainly Google search advertisements – is barely affected.

The GDPR was much anticipated and was feared to affect predominantly small businesses negatively (Kottasová, 2018; BBC, 2018). We in turn find an inverted U-shaped relationship between website size (measured in visits) and the traffic reduction due to the privacy regulation. Small websites lost traffic, but so did the largest ones, while medium websites remained unaffected and may have even grown. On the intensive margin, we observe a significant deterioration of user engagement with websites following the GDPR, both in the short and in the long term. We estimate a significant reduction in average visit duration and the number of web pages visited, as well as a measurable increase in website bounce rate – the share of website visitors that leave almost immediately after arriving on a website.

This paper makes several contributions to the fast-growing literature on the impact of GDPR in the digital market. First, we investigate how the GDPR affected web traffic and user engagement, which measure the extensive and intensive margins of internet consumption. Second, we are able to explore both short and long run effects of the GDPR. This is important, as most of the effects unfold several months after the enactment of the regulation. Third, contrary to other related studies (Peukert et al., 2020; Goldberg et al., 2019), our empirical strategy is based on the geographical source of website traffic, rather than on the country domain.[5] In such a framework, the treatment assignment targets precisely the European audience that falls under the scope of the GDPR. Finally, we assess empirically the differentiated effects of privacy regulation based on website size (Campbell et al., 2015; Dimakopoulos and Sudaric, 2018; Sabatino and Sapi, 2021). To the best of our knowledge, we are the first in providing evidence of an inverted U-shaped relationship between website size and privacy regulation.

We document and measure some anticipated and less expected effects of the GDPR.

[5] A country domain is for example *.de* for Germany, and *.fr* for France in a website's URL.

While the effects we emphasize here indicate an overall loss of traffic via most channels and less interaction with websites, this in itself does not imply that the GDPR is a harmful regulation. On the contrary: the decline of direct website traffic – the largest source of website visits – may be the result of users' conscious choice, after the increased prominence of privacy policies and websites' popups to obtain consent to these policies.

The remainder of the paper is organized as follows. Section 2 presents an overview of the related literature and our contribution. Section 3 describes the main provisions of the GDPR. Section 4 introduces the data used in our empirical analysis. Section 5 discusses our empirical strategy. Section 6 reports our main findings. Section 7 provides robustness checks. Section 8 concludes.

2 Literature Review

Our research draws from the growing body of the economic literature on data and privacy regulation, surveyed recently and extensively by Acquisti et al. (2016). This literature emphasizes the trade-off between consumer protection from data exploitation and market efficiency. In their seminal paper, Goldfarb and Tucker (2011) study the effectiveness of advertising in the EU after the implementation of the 2002 EU E-Privacy Directive, finding that the policy change reduced advertising effectiveness. Campbell et al. (2015) argue that privacy regulation imposes costs on all firms, but disproportionately on smaller ones. On the contrary, focusing on the 2009 European E-privacy Directive, Sabatino and Sapi (2021) find that mainly large firms were negatively affected, while small firms experienced no significant negative effects. Our results partially support both views: while we observe the largest traffic reduction for small websites, top websites lose traffic as well, while medium ones remain unaffected.

Our work adds to the recent line of empirical research studying the impact of the GDPR in digital markets. Gal and Aviv (2020) argue that under certain conditions the GDPR may reduce welfare by increasing data market concentration and limiting synergies. Johnson et al. (2020) find a 15% reduction in website operators' use of web technology providers following the GDPR, although this drop recovered within a few months of time. Moreover, they find that small providers were more strongly affected, and that Google and Facebook's market shares increased as a consequence of websites substituting from smaller to larger technology vendors. Peukert et al. (2020) report a decrease in third-party trackers and cookies and an increase in first-party trackers following the introduction of the GDPR. They identify a significant increase in the prominence of the leading firm, Google, at the expense of competitors.

In the same strand, Libert et al. (2018) observe significantly less third-party cookies set without consent on a sample of EU news websites in the period immediately before and after the introduction of the legislation, but Sørensen and Kosta (2019) do not find conclusive evidence of a reduction in the number of third party requests. Still in this strand, Aridor et al. (2020) investigate the effect of the GDPR on firms' ability to track users and generate revenues through online advertising. Their findings highlight a reduction in the number of consumers, but also an increase in the effectiveness of tracking on the remaining users. We contribute to this literature by providing evidence of the impact of the GDPR on websites' traffic – a measure of consumption in the digital domain – and user engagement.

Two recent papers are closely related to our research. Goldberg et al. (2019) examine the GDPR's impact on website pageviews (and revenue) for over a thousand websites, using data for 32 weeks around the GDPR's enactment from Adobe's website analytics platform. The study documents a reduction of approximately 12% in both EU user website pageviews and website e-commerce revenue after the GDPR's enforcement deadline. The reduction is larger for traffic originating from email and display adverts. The authors find no evidence that consent interfaces would dissuade users from browsing sites.

Our paper departs from Goldberg et al. (2019) in several respects. First, our data includes more than $6,500$ country-domain pairs. The large number of domains in our dataset allows us to analyze heterogenous effects by domain size category (in terms of visits). This is an important contribution, because we find significantly different effects of the privacy regulation by website size. Second, we analyze a longer time period, with our data encompassing 95 weeks, from the 5th of January 2018 to the 25th of October 2019. This is an important addition, because we document larger effects in the long-run, possibly linked to the issuance of the first large GDPR fine in early 2019. Finally, our identification strategy differs from that of Goldberg et al. (2019). The authors take as preferred control group the same set of sites in the year before GDPR enactment. Since we observe traffic by source country, we can assign treatment based on the geographic source of traffic, rather than a domain's location.[6] Despite these differences, some of our main findings are remarkably close to those of Goldberg et al. (2019). Most notably, we also observe an around 10% reduction in website visits, and a particularly large reduction of traffic originating from email marketing and display advertising.

Another closely related paper developed independently from ours is Schmitt et al. (2021). The authors analyze the impact of the GDPR on web traffic and user engagement using the same data provider as we do in this article. In line with our findings, Schmitt et al. (2021) also report an increasing long-term negative effect of the GDPR on web traffic. Our study differs from Schmitt et al. (2021) in several ways. First, we exploit additional disaggregation of the data to observe not only the country of traffic origin but also the channel that generate traffic. This is important, as we find that some paid traffic acquisition channels are particularly adversely affected. Second, we find significant long-run effects, consistent with the step up of enforcement and the issuance of the first large fine by the French regulator in January 2019. Finally, we are the first to provide evidence on a possible inverted U-shaped relationship between the traffic reduction induced by the privacy regulation and website size, as measured by the number of visits before GDPR.

There is a small but growing body of empirical literature analysing the broader economic impact of the GDPR. Jia et al. (2021) study the effect of the regulation on venture capital investment in new technology firms, finding a reduction in the number of monthly deals in the EU compared to the US. Zhuo et al. (2021) study the effect of the GDPR on interconnections among network operators. The authors report that the legislation had no impact on the number and type of agreements at the internet layer and may have had a minor effect on the entry and the number of network customers. Through an experiment with a large telecom

[6] Since the GDPR applies to EU residents, regardless where the website they visit is located, our identification strategy is fully in line with the logic of the GDPR.

provider, Godinho de Matos and Adjerid (2021) find evidence suggesting that more informed final user consent to data processing mandated by the GDPR can have positive effects to both consumers and firms.

3 Institutional Background

The General Data Protection Regulation (GDPR) was passed in April 2016 and came into effect on the 25th of May 2018 across the EU. The regulation is labelled *"general"* because it applies to any firm handling data, offline and online likewise. The regulation was well anticipated, and the market had a two-year period between enactment and entry-into-force to prepare. The GDPR aims to strengthen and harmonizing the legislation concerning privacy, data collection and processing within the European Union. It reshapes the way personal data of EU residents are collected and processed, providing new rights to access and control these data. These improved user rights relate mainly to access, rectification, erasure, and portability of user data. The GDPR places the burden of justifying data processing with lawful motivation upon data controllers, who are obliged to obtain explicit and informed consent from data subjects in order to handle those data. The legislation imposes obligations on organizations operating anywhere in the world as long as they collect or process EU residents' data. Furthermore, it imposes large fines for infractions, up to 20 million euro or 4% of worldwide annual revenues, whichever is higher.[7]

The GDPR can impact website traffic by reducing the intensity and effectiveness of online advertising in (at least) three ways. First, by interfering with tracking technologies such as cookies, trackers, fingerprinting, beacons – which are used to gather data on users' activities throughout the Internet.[8] Typically, such tracking happens without the user being aware of it (The Economist, 2014). The GDPR affects this mechanism by limiting data

[7] The global scope of the legislation and the severity of the fines are key differences compared to Directive 95/46/EC, which the GDPR replaced.

[8] Through the collection and processing of these data, the operator – either the website or a third-party web technology vendor – can infer user preferences: websites can exchange and pool visitor identifiers and visit logs. By combining these logs and linking them to individual visitors, the online advertising industry is able to create detailed user profiles with visit histories, interests and purchases. This information can then be used for targeting advertisements to the user, when she visits an ad-supported webpage.

processing to well-specified cases.[9] Moreover, a website using web technologies to identify and target users can do so only after obtaining explicit and informed consent.[10] Restrictions in data collection and processing can reduce both the effectiveness and the intensity of advertising (Goldfarb and Tucker, 2011). The intensity can be reduced because of higher costs incurred by firms, or because some forms of advertising become unlawful. For example, email advertising may decrease because email lists might have been redacted as a result of the GDPR since email addresses qualify as data protected by the regulation. Therefore, firms cannot send unsolicited emails to any address they collected. On the other hand, the effectiveness of advertisement may decrease because, as fewer users consent to data collection and processing, advertisement becomes less tailored to the specific user.

The second way through which the GDPR can impact online advertising is by affecting the perceived legal risks of website operators. Under the GDPR, websites are jointly responsible with third-party providers. As a consequence, both may reduce advertisement to avoid the risk of sanctions. Peukert et al. (2020) find evidence that fear of legal repercussions led to an increase in concentration in the website technology vendor market, as website operators perceived top players as more trustworthy.

Finally, the GDPR can impact advertising by affecting online user behavior. From its approval to its enforcement, the legislation has been widely discussed, bringing privacy concerns to the forefront of the public debate.[11] The boom in online pop-ups asking for consent to place cookies on the user's device has been hard to miss. European internet users will also long remember the storm of emails from various websites asking to opt-in to various mailing lists before and shortly after the enactment of the GDPR. The greater awareness,

[9] Specifically, the fulfilment of a legal obligation or a contract, the protection of vital interests of a person, reasons of public interest, firm or a third party's legitimate interest and, finally, consumer consent. Whether online tracking does or does not fall into these motivations is debated. At the very least however, the GDPR increased the risk for firms to wrongly classify their practices into these bins, since they are liable and are subject to fines.

[10] The GDPR is not the first regulation to introduce the prerequisite of user explicit consent (opt-in) for data collection. Already under the revised 2009 E-Privacy Directive, consent was needed to place cookies on a users' device. Thus, the key difference of the GDPR has been to extend the scope of the territorial applicability, to sensibly increase fines in case of infringements, and to adopt a technology-neutral approach.

[11] A search on Google Trends on the term "*GDPR*" illustrates this well: there was a large surge in English language web-searches on this keyword around May 2018, the entry into force of the regulation. https://trends.google.com/trends/explore?date=2018-01-01%202019-12-31&geo=GB&q=gdpr.

in turn, can influence the degree to which users actively try to protect their privacy. User behaviour can impact advertising by avoiding websites that are recognized as more intrusive, choosing not to opt-in to data handling policies, or increasingly using technologies that limit tracking and advertisement.

From a legal perspective, there has been little enforcement following the entry into force of the GDPR. In the second half of 2018, only nine fines for GDPR infringements were imposed around Europe, for a total of only $440,000$ euro. The first significant fine – the third largest to date – came on the 21st of January 2019. On that occasion, the French regulator imposed a 50 million euro penalty on Google "*for lack of transparency, inadequate information and lack of valid consent regarding the ads personalization*" (CNIL, 2019). That fine was perceived by some commentators as the end of an unofficial grace period, with The Economist going as far as calling it "*the start of a war*" (The Economist, 2019). Figure 1 shows the cumulated fines from July 2018 to October 2019. The jump in January 2019 coincides with the French Google penalty. In retrospect, the following months show no increase in regulatory enforcement that would have fulfilled the expectations of the time. However, the prominence of the first large fine – issued to a company as visible as Google – might have influenced firms' compliance decisions.

4 Data

4.1 Data Description and Summary Statistics

Our main source of data is SimilarWeb, a web analytics service provider.[12] Similarweb tracks information on website traffic volume – the total number of visits for each domain – as well as several measures of user online behavior. The provider records daily data along the following key metrics:

- **Visits**: Total visits to a domain via all traffic channels. A visit is the access of a user to one or more pages of a website. All subsequent page views belong to the same visit,

[12] www.similarweb.com

until the visitor is inactive for 30 minutes.[13]

- **Average visit duration**: The average visit length measured in seconds, which is calculated as the time elapsed between the first and last page view of a visit.
- **Bounce rate**: The share of users who close the website after landing on it without loading other pages.
- **Average pages per visit**: The average number of pages viewed per visit.

Visits is our main outcome of interest. We refer to the other three outcomes as engagement measures, as they provide information on the way users interact with the website.

A website is defined by its domain, so we use both terms interchangeably.[14] A domain can receive traffic from different countries and through different *channels*. For instance, a visitor from the UK can land on an e-commerce website registered anywhere in the world through an advertisement seen on a social platform. Our data allow us to observe the visits by country of origin of the visitor, as well as the traffic channel through which the website was reached. In particular, Similarweb categorizes traffic into the following channels:[15]

- **Direct**: Website visits by users directly typing the website address into the URL bar in a browser. Direct traffic also covers clicking on a bookmark or on a link from outside the browser (but not in emails).
- **Display Advertising**: Traffic from users clicking on an advertising banner or video advertisement shown on a third-party website (except social networks).
- **Email**: Website visits following the user clicking on a link provided by a web-based mail client. For example, clicks in marketing emails on links pointing to the website.
- **Organic search**: Traffic originating from the organic (i.e. non-paid) results in an internet search engine such as Google or Bing.

[13] We use the terms visits and traffic interchangeably.

[14] This means that we treat google.com and google.fr as different domains. While a domain like sites.google.com may host several websites, they constitute a single domain for our analysis.

[15] Note that these are the same traffic channels Google Analytics tracks. They can be therefore seen as the industry standard classification.

- **Paid search**: Traffic originating from clicks on paid advertisements on the result page of a search engine, such as Google or Bing.
- **Referral**: Traffic generated by third-party links on most other websites than social networks and search engines. This includes paid links in blogs and other affiliates, and free traffic such as media coverage.
- **Social**: Traffic coming from social media websites such as Facebook or YouTube, either through users posting links for free or advertisement.

Finally, for every domain Similarweb also reports the share of traffic originating from each country at a monthly level, along with domain's traffic rank in that country.

We select the largest websites active in the US and in major European countries. In particular, we observe the top $1,000$ domains in five EU countries: France, Germany, Italy, Netherlands and United Kingdom, together with the top $2,000$ US domains in terms of visits. Since domains are ranked by traffic volume within each country, we collect $7,000$ domain-country pairs. However, as some domains belong to more than one country's top thousand (e.g. amazon.com), we have data for $5,300$ unique domains. Of these, we keep the $4,957$ websites that were active both before and after GDPR.

Traffic data is further disaggregated by channel, so that the cross-section is defined by the triple domain-country-channel and comes in daily frequency.[16] Our data span two years from 01-11-2017 to 31-10-2019, covering about seven months before and 17 months after the GDPR's implementation. The panel is unbalanced, as some domain-country couples have missing data for various periods, with no clear pattern.[17] Since traffic data show a high degree of daily volatility, we aggregate the observations at the weekly level. We define a week as a Friday-to-Friday period, as the GDPR entered into force on Friday, 25th of May 2018. The panel covers 94 weeks, from the 5th of January 2017 to the 25th of October 2019. The final dataset covers around 4 million observations at the channel-domain-country-week level. Summary statistics are provided in Table 1.

[16] An observation in our dataset for example corresponds to the number of daily visits to bbc.com in the United Kingdom via clicks on organic search results.

[17] Missing observations do not happen systematically over time or across websites.

We consider our list of websites very extensive. The websites in our dataset cover a very significant share of websites typical users in the selected countries ever visited. Website traffic appears highly concentrated, as reflected in our data: Figure 2 shows the average weekly traffic by percentile of the distribution of domain's worldwide traffic before the GDPR.[18] The traffic for every percentile is shown on the left panel, while the right panel removes the top 1st percentile to avoid flattening the scale. It is evident that the distribution is highly skewed towards the top: The handful of most visited websites attract the vast majority of internet visits. Therefore, selecting the top $1,000$ domain by country ($2,000$ for the US) ensures that our sample covers a high share of internet traffic in these countries and can be regarded as representative.

Our data by channel allows us to distinguish between paid and unpaid traffic. The former generates either a remuneration to a third party (e.g. advertisement fees for Google) or is the result of an internal marketing strategy (e.g. email campaign, where email addresses may be purchased). Direct traffic is generated by directly typing a domain's URL into the browser, or clicking on a bookmark. In this view, we consider Direct traffic and Organic Search as unpaid traffic, while the other channels are labelled as paid.

Figure 3 shows the traffic distributed by channel. The left panel reports the contribution of each channel on total website traffic. Most of the traffic comes from the Direct channel, amounting to about 67% of total website traffic. Organic search makes up for another 20%. Taken together, these unpaid channels amount to almost 90% of overall traffic. The remaining 12% of overall website traffic can be regarded as *paid*. The right panel of Figure 3 shows the channel shares for paid traffic only. While referrals and social networks constitute the bulk of paid traffic, each channel has a non-negligible share, with email amounting to 20% of website paid traffic.

[18] We obtain this distribution by taking each domain's weekly worldwide traffic (i.e., not limited to the six countries in our analysis) and averaging it across the time periods that precede the introduction of the GDPR. Then, for each percentile, we calculate the average weekly traffic depicted in Figure 2.

4.2 Recording Bias

A potential concern affecting our data is that the GDPR may have had an impact on the ability of our data provider to measure website traffic and usage. (Goldberg et al., 2019) refer to this as the *recording bias*. For example, if users increasingly opt-out of any form of data sharing with third-parties, websites may have been less able to report data about these metrics. We took several steps to investigate to what extent the recording bias may arise on our data.

First, we explicitly inquired with Similarweb about a potential bias in measurement associated with the GDPR. Similarweb confirmed in writing that this was not the case.

Second, according to publicly available information and Similarweb's own account, the data provider combines data from several sources, including websites directly, internet providers, public data sources as well as an anonymized panel of users of one or more browser extensions.[19] Much of these data are not individually identifiable, and therefore largely fall outside the GDPR.[20] No user-level data is necessary for any of the analysis we do. Similarweb also declares publicly not relying on any personally identifiable data in its methods, again, largely rendering it immune to measurement issues due to the GDPR.[21] In particular, it explains that: *"We employ a multi-step verification process to ensure data collected is devoid of any Personally Identifiable Information (PII) [...] Behavioral data is shared anonymously and aggregated at the site- and app-level rather than the user-level [...] Data is never used for advertising or targeting, and we don't use "cookies" to collect behavioral data."*

Third, we also gather additional information on Similarweb's user panel, which is partly based on tracking users' browsing behaviour through its own browser extension.[22] Similarweb

[19] https://licreativetechnologies.com/seo/how-to-track-website-traffic-similarweb/, retrieved on the 1st of February 2022.

[20] This even applies to such seemingly sensitive items as gender, which we do not use in our analysis: https://support.similarweb.com/hc/en-us/articles/360001253797-Website-Demographic-Data, retrieved on the 1st of February 2022.

[21] https://support.similarweb.com/hc/en-us/articles/360001631538-Similarweb-Data-Methodology, retrieved on the 1st of February 2022.

[22] See for example https://chrome.google.com/webstore/detail/similarweb-traffic-rank-w/hoklmmgfnpapgjgcpechhaamimifchmp, retrieved on the 2nd of February 2022.

makes users of the web extension aware of being tracked, and in exchange it provides statistics on the websites they visit. In February 2022, the add-on had more than $800,000$ users in the Chrome web store alone, not counting users of other browsers. Given that we focus on the top few thousand websites in the largest European countries and the United States, it appears to us that solely based on such a panel Similarweb would be able to capture the metrics we analyze in great detail. As the Chrome extension was already available on the 10th of August 2017, it appears unlikely that users of such a browser extension would have revoked their consent to be tracked due to the GDPR.[23]

Finally, Similarweb is a widely audited industry-standard source of website traffic information, providing data to several companies engaged deeply in web-traffic measurement, such as Adobe, Google and The Economist, as well as to researchers ((Calzada and Gil, 2020; Lu et al., 2020; Schmitt et al., 2021)). Schmitt et al. (2021) compare Similarweb's data with a German data provider on web audience (AGOF) which was likely not affected by GDPR, finding no significant deviation between the two sources.

Overall, we conclude that our data are fit for purpose and unlikely to suffer from the recording bias. We now turn to the results of our main analysis.

5 Empirical Model

Our main interest is understanding how the GDPR affected website traffic as well as user behavior on the internet, in the short and in the long term. Our empirical strategy applies a difference-in-differences (DiD) approach to compare our outcomes of interest in EU countries with the US, before and after the introduction of the GDPR.

Our data are unique as they allow us to identify the geographic origin of a website's audience, as opposed to the country of registration of the domain. Our treatment assignment reflects the main provision of the GDPR, since it is defined as the traffic coming from one of the EU countries that fall under the GDPR. Therefore, our definition of treatment and control

[23] We found the first historic version of the web page from the *Wayback Machine*, https://web.archive.org/web/20170810122245/https://chrome.google.com/webstore/detail/similarweb-traffic-rank-w/hoklmmgfnpapgjgcpechhaamimifchmp

groups is very precise: it does not derive from the domain of the observed website, but rather on the geographical source of traffic for the specific website. As a consequence, each website can be partially treated if its traffic comes from both the US and EU countries. Another important dimension that we exploit refers to the different types of traffic channel. That is, for a specific geographical source of traffic, we observe the channel through which the website is reached. This further increases the sources of variation in our analysis, and allows us to investigate potential heterogeneous effects of the GDPR on specific traffic channels.

Our baseline model takes the following form:

$$y_{i,k,c,t} = \beta_0 + \beta_1 Post_t + \beta_2 EU_c + \beta_3 Post_t \times EU_c + \alpha_i + \kappa_k + \epsilon_{i,k,c,t} \tag{1}$$

where $y_{i,k,c,t}$ is one of our outcomes of interest[24] for website i, from channel k and country c at time (weekly date) t.[25] *Post* is an indicator identifying the period from the introduction of the GDPR onward (the 25th of May, 2018). *EU* is a dummy identifying whether the country from which website traffic comes is an EU member state. α and κ are website and channel fixed effects respectively. Finally, ϵ is a mean-zero error term.

The main coefficient of interest in Equation (1) is β_3, which captures the average causal impact of the GDPR on website outcomes across channels. For example, taking website visits as the dependent variable, the estimated coefficient measures the average (percentage) difference in traffic coming from the EU and the US induced by GDPR.[26] Since we also want to assess the main channels through which the GDPR affects web traffic, we further interact *Post*, *EU* $Post \times EU$ with a full set of dummies identifying traffic channel k. In this setting, we estimate differentiated causal effects of the GDPR for each traffic channel.

Equation (1) does not allow us to disentangle short and long term impacts of the GDPR. In particular, if website operators anticipated weak enforcement of the regulation during the

[24] Our dependent variables are *visits*, *average visit duration*, *pages visited*, and *bounce rate*. For the first three variables we apply the following monotonic increasing transformation $\ln(x + 1)$, so as to include null values in the estimation. Since the bounce rate takes values between zero and one, we do not apply any transformation to it.

[25] As explained in Section 4, the traffic acquisition channels are *Direct*, *Display Avertising*, *Paid* and *Organic Search*, *Referrals*, *Social* and *email*.

[26] The logic is analogous when the depend variable is one of our engagement measures.

subsequent months following the GDPR's introduction, then we might expect to see little effect during such a period. In this case, β_3 would underestimate the true effect of the regulation. For this reason we also run the following alternative model

$$y_{i,k,c,t} = \beta_0 + \beta_1^S Post_t^S + \beta_1^L Post_t^L +$$

$$+ \beta_2 EU_c + \beta_3^S Post_t^S \times EU_c + \beta_3^L Post_t^L \times EU_c + \alpha_i + \kappa_k + \epsilon_{i,k,c,t} \tag{2}$$

where $Post^S$ identifies the period between the 25th of May 2018 and the week before the 21st of January 2019, that is the period between the entry into force of GDPR and Google's fine in France, while $Post^L$ is active from the end of January 2019 onwards. Thus, in this specification the coefficient β_3^S identifies the short-run causal impact of the GDPR, while β_3^L identifies the long-run effect on the outcome variables. Additional interactions of $Post^S$, $Post^L$, EU, $Post^S \times EU$, and $Post^L \times EU$ with channel-specific dummies allow us to identify diverse effects across traffic channels both in the short and in the long term.

Finally, to understand the dynamics behind the estimated parameters of interest in both (1) and (2), we interact our treatment (EU) dummy with time specific dummies, leading to the following Equation:

$$y_{i,k,c,t} = \beta_0 + \sum_t \beta_{1,t} Time_t \times EU_c + \alpha_i + \tau_t + \gamma_c + \kappa_k + \epsilon_{i,k,c,t} \tag{3}$$

where τ and γ denote time and country fixed effects respectively. We also estimate Equation (3) by traffic channel in order to assess potential differential dynamics across those channels.

5.1 Parallel Trends Test

Equations (1) and (2) identify the causal impact of the GDPR on web traffic and user engagement if US online behavior provides a valid counterfactual for EU countries. The identifying assumption is that, absent GDPR, EU and US traffic would have had a similar pattern. We test such an assumption by checking for parallel trends before GDPR enactment.

Conditional on fixed effects, we should observe a similar trend in the dependent variables between treatment and control pre-GDPR.

We start by estimating the following Equation separately for the US and EU:

$$y_{i,k,c,t} = \beta_0 + \tau_t + \alpha_i + \kappa_k + \epsilon_{i,k,c,t}. \tag{4}$$

We then plot the estimated coefficients associated to the time dummies τ_t for both the US and EU in Figure 4, together with the corresponding 95% confidence interval. In doing so, we exclude the first-period dummy to avoid multicollinearity. Hence, each coefficient represents a (percentage) variation of the dependent variable normalized to the first week. Each panel corresponds to one of our variables of interest, namely web traffic (Panel a), average visit duration (Panel b), average pages visited (Panel c), and bounce rate (Panel d).

It is reassuring to observe that, before GDPR, the US and Europe behave similarly for all dependent variables, suggesting that the US provides a valid counterfactual for European online behavior. What is more, US and EU follow a similar path after GDPR, reacting very similarly to seasonal shocks.[27] The picture also shows a significant reduction in European traffic with respect to the US, both in the short and in the long term, with a similar pattern emerging also for average visit duration. On the contrary, as for average pages visited and bounce rate, we only observe a long-run separation of the two trends. These preliminary pieces of evidence suggest that (i) the US and Europe behaved similarly before GDPR, and (ii) GDPR negatively affected web traffic and user engagement in European countries.

We also test formally for parallel linear trends before the GDPR between our treatment and control groups by estimating the following Equation:

$$y_{i,k,c,t} = \beta_0 + \gamma_1 Trend_{EU} + \gamma_2 Trend_{US} + \alpha_i + \kappa_k + \epsilon_{i,k,c,t}, \tag{5}$$

where $Trend_{EU}$ and $Trend_{US}$ are European-specific and US-specific linear trends respec-

[27] For instance, looking at Figure 4 Panel (a), we observe a drop in web traffic few weeks before and after Google's fine both in the EU and US. The same happens for the engagement metrics, where the two curves react in parallel over time.

tively. We estimate Equation (5) only for the period before GDPR. Ideally, one would like to observe similar estimates for γ_1 and γ_2, implying no differential trends between treatment and control before the shock.

Table 2 shows the estimated coefficients of Equation (5) for each variable of interest. Overall, we observe similar trends between the US and Europe, particularly for the engagement measures (Columns 2-4). In fact, when we test for the equality of the two coefficients of interest, we cannot reject the null hypothesis for average visit duration, average pages visited and bounce rate. On the contrary, we do reject that the two trends are equal for web traffic at the 5% level (Column 1). However, the difference between the two coefficients is fairly small (0.0016), most likely determined by the minor jump a few weeks before the GDPR that we observe in Figure 4 Panel (a).

In summary, our parallel trends tests suggest that US traffic provides a valid control group for European traffic. When we analyze non-parametrically both trends, we find no systematic deviations before GDPR. When we run a formal test of equal linear trends, we reject the null hypothesis of equal trends only for traffic. However, both the magnitude of the difference in trends and the F-test are small, suggesting the absence of a significant divergence in trends between treatment and control group.

6 Results

We start with the results on website visits, looking first at the overall impact of the GDPR on traffic and then analysing effects by traffic channel and website size category. We then discuss the results for visitor engagement.

6.1 Website Traffic: Average Effect

Table 3 reports *Ordinary Least Square* (OLS) estimated coefficients for different specifications of Equation (1) when the dependent variable is the natural logarithm of website traffic (plus one). The first column collects estimates from our baseline model. The coefficient associated to *Post* is positive and statistically significant, suggesting a general expansion in web traffic over time. At the opposite end, the coefficient associated to EU is strongly negative, implying that European countries experience lower traffic compared to the US. The coefficient associated to $Post \times EU$ captures the impact of the GDPR on web traffic. The coefficient is negative and strongly significant, and points to a reduction in web traffic of about 10%.[28] Hence, following the introduction of the GDPR, European website traffic experienced a significant reduction compared to US traffic.

Columns (2)-(6) report estimated coefficients of Equation (1) as we add different sets of fixed effects. It is interesting to observe that $Post \times EU$ coefficient barely moves, which increases our confidence in the baseline specification. Finally, in Column (7) we also control non-parametrically for website-specific time trends by including domain-time fixed effects. This specification identifies the main parameter of interest by exploiting time-variation across country of origin within each domain, which implies that only domains receiving traffic from more than one country contribute to the estimation of $Post \times EU$. That is, the estimated coefficient captures the causal impact of the GDPR only for websites with multinational traffic. The coefficient is still negative and strongly significant, although smaller in magnitude. Since multinational domains are typically larger websites, at this point we might expect potential heterogeneous effects of the GDPR on web traffic based on website size (Section 6.3).

To identify long and short term effects of the GDPR on web traffic, we iterate the same battery of regressions on Equation (2), where we split the $Post$ dummy into $Post^S$ and $Post^L$, which identify short and long run periods from the GDPR introduction respectively. In particular, $Post^S$ is active from the GDPR's enactment (25th of May 2018) until the week before the French fine imposed on Google – the so-called "*start of war*" by data protection authorities against privacy violators (The Economist, 2019) – which happened on the 21st of January 2019. On the other hand, $Post^L$ takes value one from that date onward. Hence, the interactions $Post^S \times EU$ and $Post^L \times EU$ identify the causal impact of the GDPR on web traffic in the short and in the long term respectively.

Table 4 collects estimated coefficients of Equation (2). From Column (1) we observe

[28] This is fully in line with the estimated traffic drop reported by Goldberg et al. (2019).

that both $Post^S$ and $Post^L$ are positive and significant, with the latter being larger in magnitude. The coefficients associated with $Post^S \times EU$ and $Post^L \times EU$ are both negative and significant. Moreover, they are statistically different from each other. The estimated $Post^S \times EU$ coefficient implies a reduction of 4% of web traffic induced by the policy change in the eight months after the introduction of the GDPR, while the $Post^L \times EU$ estimated coefficient suggests a much larger effect in the long-run of -15.7%. Again, estimates barely move when we add different sets of fixed effects (Columns 2-6). When we exploit only multinational website heterogeneity, the coefficients follow a similar path, although both short and long term effects are lower in magnitude. Thus, results from Table 4 suggest that most of the negative effect on web traffic materializes in the long-run, after January 2019. As a consequence, the timing of enforcement might have played a major role in spurring compliance with the law and, therefore, on the variation in website traffic.

What are the dynamics behind the aforementioned results? One possibility is that DiD estimates may hide potential time-varying treatment effects. That is, the impact of the GDPR on web traffic increases over time, implying a divergent path in the post-GDPR period between EU and US traffic. Another possibility is the existence of a transitory period that may coincide with the lack of enforcement and compliance characterizing a *grace period* in the months right after GDPR enactment. In such a case, a potential impact on web traffic may take some time to materialize. We investigate this issue by estimating Equation (3), in which we interact the EU identifier with time-specific dummies after controlling for domain, channel, time, and country fixed effects.

Figure 5 displays the estimated coefficients (hard black line) and 95% confidence intervals (dashed lines) from the dynamic DiD of Equation (3). Red vertical lines identify the entry into force of the GDPR and the end of the grace period identified by the French financial penalty imposed to Google. First, we notice that pre-GDPR estimated coefficients are fairly constant, and almost all of them are not statistically different from zero. This is in line with our parallel trends test, as it suggests no diversified trends between treatment and control groups before the policy change. Second, we observe a negative trend in EU web traffic *vis-à-vis* the US starting after the 25th of May 2018 – the date of GDPR enactment –

which stops after Google's fine. Third, estimated coefficients stabilize in the long-run, implying an average 15% reduction in web traffic compared to the pre-GDPR period. Hence, the dynamic DiD highlights that the downward shift in the EU web traffic does not materialize right after the introduction of the GDPR. Consistently with the idea of a strict relation between enforcement and compliance, the negative impact of the GDPR on EU web traffic arises eight months later the regulation's entry into force.

In summary, our DiD estimates suggest a negative effect of the GDPR on web traffic. The effect is large and economically significant, implying an average 15% reduction in long-run web traffic. However, this effect does not materialize immediately after the entry into force of the GDPR, but unfolds fully the following year. The dynamic DiD highlights that the transition started during the grace period and stopped a few weeks after the end of the regulatory holiday period, emphasizing the role of enforcement on website compliance with GDPR rules.

6.2 Analysis by Traffic Acquisition Channel

How did the GDPR affect website traffic by various traffic channels? If the effect by such channel differs, is the loss of a specific channel counterbalanced by an increase in traffic in another channel? We answer these questions by interacting $Post^S$,$Post^L$, $Post^S \times EU$, and $Post^L \times EU$ from Equation (2) with channel-specific dummies in order to identify diversified effects across traffic channels both in the short and in the long term.

Table 5 collects estimated coefficients $Post^S \times EU$ and $Post^L \times EU$ for each traffic channel. First, we notice that the short-run effects are smaller in magnitude compared to long-run ones for every channel. We observe a short-run negative impact of the GDPR in selected channels, including Direct, Email, Referrals, and Social. The largest coefficient is for Email traffic, suggesting a reduction of about 8% in this channel. A smaller but strongly significant effect arises for Direct traffic, where the coefficient implies a reduction of about 4%. Hence, also unpaid traffic channels have been affected by GDPR restriction, and in the same direction as paid channels.[29]

[29] This is consistent with the view that different traffic acquisition channels are complements, rather than

Most of the negative effect of the GDPR on web traffic materializes in the long-run, specifically after the grace period at the beginning of 2019. This is true for each traffic channel, as shown by the channel-specific $Post^L \times EU$ coefficients in Table 5. The impact on the Email and Display advertising channels is particularly severe, as the coefficients imply a reduction in traffic induced by the GDPR of about 35% and 29% respectively.[30] Although these coefficient estimates are rather large for email and display advertising traffic, they also reflect the relatively small amount of traffic coming from those specific channels. In fact, on average less than 3% of website traffic comes from Email and Display ads (Figure 3), implying that even a small absolute reduction in traffic may account for a large variation in percentage points.

We also observe a significant reduction in traffic coming from other paid channels, such as Referrals and Social. Only the coefficient associated to Paid Search is not statistically significant at 5%, neither in the short nor in the long term.[31] What is more, unpaid channels are negatively affected too. Our results suggest a reduction in Direct traffic of about 9%, while traffic from Organic Search experience a 7% reduction. This is somewhat surprising because we would not expect these channels to rely on targeted messages to the same extent as paid advertising does. The effect likely comes through increased user awareness about privacy: the GDPR triggered an increased use of various information banners where websites seek user consent to data handling upon entry. A not insignificant group of users likely turn away from websites upon being presented these popups.

We find that both paid and unpaid website traffic channels are negatively affected by the GDPR in the long-run. Paid channels are more severely affected, although also Organic Search and Direct traffic reduce. Finally, the coefficients in Column (6) of Table 5 suggest that the impact of the GDPR on traffic is significantly smaller for multinational websites.

substitutes.

[30] The strong negative effect on email traffic is in line with industry testimonies about the "*death of email marketing*" (Harris, 2018).

[31] This is as expected. Personal data is not used much in paid search. As explained by a recent report of the UK competition authority: *"search ads rely only in a limited way on personalisation, rather they are primarily targeted to match key search terms entered on search engines (ie the 'search query'), which typically provides most of the information needed to serve a relevant ad"*, https://assets.publishing.service.gov.uk/media/61b86aee8fa8f5037ffaa347/Appendix_I_-_Considering_the_impacts_of_Apples_ATT.pdf, retrieved on the 1st of February 2022.

In the long-run only paid channels such as Email, Display Ads, and Social are negatively affected, with the magnitude being significantly lower compared to the ones in Columns (1)-(5). Hence, potential heterogeneity may arise depending on firm size, paving the way to our dedicated analysis in Section 6.3.

6.3 Analysis by Website Size

We now turn to the analysis of the effect of the GDPR on traffic by website size category. Figure 2 shows average visits (in thousands) for each percentile of the distribution of worldwide traffic before the GDPR. It highlights how website traffic is distributed in a highly heterogeneous way, with the presence of a sizeable group (59%) of small websites that however account only for a small share of total traffic (6%), a modest group (30%) of medium ones (13% of traffic), few (10%) large websites (23% of traffic) and a handful (1%) of giant websites that account for the largest share of traffic (57%). Due to the significant heterogeneity in website size, we analyze whether the effects of the GDPR differ for different size domains. We take into account the high skewness of the traffic distribution when classifying websites by size. We split domains in four categories according to the percentile of worldwide traffic they fall in: small ($1 \leq p \leq 59$), medium ($60 \leq p \leq 89$), large ($90 \leq p \leq 99$), giant ($p = 100$)

Table 6 reports estimated coefficients for two different specifications of Equation 2 by website size. We observe that the short-run effect ($Post^S \times EU$) is negative and statistically significant only for small websites, implying a traffic reduction of about 14%. The long term effect ($Post^L \times EU$) is negative and statistically significant for all websites but medium ones. The coefficient varies strongly by website size, implying a reduction of website traffic of about 41% for small websites, and 7% and 16% for large and very large websites, respectively. These results are robust to different specifications, with coefficients remaining practically unchanged when adding interacted fixed effects.

Overall, our results point to an inverted U-shaped long-run relationship between website size and traffic change due to privacy regulation: The smallest and the largest websites registered a significant drop in visits, while medium ones lost fewer visitors. This is an important result that nuances policy discussions preceding the GDPR, where commentators warned about small firms being particularly severely affected by privacy regulation (Cherry, 2017). The European Commission called the allegation that "*GDPR is overwhelming for small businesses*" a "*myth*" (European Commission, 2019). Our results show that small websites were indeed hit particularly hard by the GDPR. However, so were large websites as well. Medium websites in turn remained largely spared from associated traffic loss.

6.4 User Engagement

Results from Tables 3-4 suggest a reduction in website traffic due to the GDPR's enforcement. That is, websites receive less visits from users. However, the restrictions imposed by the GDPR may affect also user behavior once the visitor has effectively reached the website. Hence, the policy change can affect both the extensive (website visits) and intensive margin (website engagement) of online behavior. For instance, a less effective display ad may drive the user to a dull website for which she has no particular interest. In such a situation, we can expect the user to quickly shift to another activity, either offline or online.

We investigate whether the GDPR has affected also the intensive margin of online behavior by estimating Equation 2 on three metrics of user engagement, namely average visit duration, average pages visited, and bounce rate.[32] Table 7 collects OLS coefficients from such regressions. Focusing on average visit duration (first column) we observe a significant reduction both in the short and in the long run of 1.7% and 3.5% respectively. A similar pattern emerges for the average number of pages visited within the website, with estimates pointing to a reduction of 0.7% and 1.5% in the short and long term respectively. As for the bounce rate, results point toward a reduction in engagement only in the long-run of 0.006 percentage points. Thus, the GDPR has a negative impact on both intensive and extensive margins. Consistently with previous results on web traffic, the adverse effect on user engagement induced by the GDPR materializes mostly in the long-run.

Are the results driven by specific channels? Table 8 suggests that is the case. When we disentangle heterogeneous effects by traffic channel, we find that the negative impact of

[32] For a description of the three variables see Section 4.

the GDPR on user engagement comes from those users reaching the website mainly through email, although also engagement from social networks and organic search traffic is significantly impacted both in the short and long term. In the long-run, average visit duration decreases mainly for paid traffic channels, including Email, Referrals, Paid Search, and Social, thus indicating online advertising being less effective also in terms of engagement.

All in all this battery of results highlights another adverse effect in addition to the drop in web traffic. User engagement is negatively affected, and particularly following the effective enforcement of the GDPR after the issuance of the first large fine. The policy change affected disproportionately paid channels, although we find also evidence of a negative effect from organic search traffic engagement, pointing toward potential complementarities across channels also in the intensive margin.

7 Robustness Checks

We run two series of robustness checks that consider the presence of potential spillovers. The first one deals with the presence of multinational traffic, which might generate spillovers between treatment and control if multinational domains apply a uniform privacy policy to both US and EU visitors. The second robustness check assesses the presence of differential effects for websites located in the EU compared to those in the US, which may also suggest potential spillovers. This might happen if EU websites comply with the GDPR also when serving US visitors, or if US websites do not comply with the GDPR at all. Both robustness checks confirm our main results, namely a reduction in web traffic and user engagement following the entry into force of the GDPR, and in particular in the long-run.

7.1 Excluding Multinational Traffic

In Equations (1) and (2) the treatment assignment covers traffic originating from the EU. Since a website can receive visitors from more than one country, this implies that multinational websites are treated only for the portion of traffic coming from Europe. This is in line with the main provision of the GDPR, which aims to protect all EU citizens regardless the website's actual location. However, an implicit assumption for the identification of the causal impact of the GDPR on web traffic is that multinational domains comply with the GDPR only for EU visitors. That is, websites are assumed to apply a different privacy policy for EU visitors and to visits coming from the US. Although we do informally observe websites treating visitors from the US and the EU differently, this may not always be the case.[33] It may be possible that websites with a high share of EU traffic comply *tout court*, while domains with most of the traffic coming from the US may prefer not to comply with the limitations imposed by the GDPR. In the extreme case, websites may decide to block access to EU visitors altogether.[34]

We test for this issue by focusing only on websites that receive most of the traffic from one of the countries of our analysis. In particular, we focus on domains with more than 85% of their global traffic coming from either one of the EU countries we follow (e.g., France, Germany, Italy, Netherlands, and the UK) or the US before the GDPR.[35] Moreover, we select only the traffic originated from the country that meets the 85% threshold. In this way, we clean the dataset from multinational domains (such as facebook.com), as well as from country traffic that cannot be clearly allocated between treatment and control group.

Table 9 collects the estimated coefficients of Equation (2) in this setting. By looking at website traffic (Column 1), we find that both the short and long term (negative) effects of the GDPR remain intact. The estimated coefficients are very much in line with our main results (Table 4 Column 5), suggesting an increasing long-run effect, with the magnitudes also being similar. The short-run effect remains largely unchanged (0.3% difference), while the long-run effect increases slightly by 2%. This suggests that increasing the precision of the treatment assignment rises the estimated impact, as one would expect. However, it also highlights that the potential bias is fairly small, which in turn increases our confidence in

[33] We casually surfed several multinational websites using a VPN service that allows us to manually set our IP address as US or European. A large number of websites provide a different experience tailored to the perceived country of origin.

[34] There is some evidence that US news websites blocked EU users right after the enactment of GDPR, because they were not able to comply with the new data protection rules in the short term. See https://digiday.com/media/u-s-sites-continue-block-european-visitors-post-gdpr/.

[35] Despite the focus on specific EU countries, Similarweb provides metric for global traffic for each domain. Given a domain's global traffic, we can derive the share of traffic coming from each of the observed countries at each point in time and for each domain.

the research design.

A similar argument applies for the engagement measures. A comparison with Table 7 shows that, if anything, cleaning from multinational traffic increases the estimated coefficients. In this case, the difference between the two estimates is significantly larger. For instance, the long-run impact on average visit duration moves from -3.5% to -5.3%, which implies a 50% increase in magnitude. Moreover, while results from Table 7 Column 3 do not suggest any significant short-term effect on the bounce rate, we now observe a short term increase of about 0.4 percentage points.

In conclusion, the results from this robustness check confirm our main findings, namely a negative impact of GDPR on web traffic and user engagement. Although the bias from multinational traffic does not significantly affect estimates on web traffic, it somewhat underestimates the impact on user engagement metrics, implying that results from Table 7 are likely even conservative.

7.2 US vs. EU Websites

Another potential source of concern might arise if websites apply a uniform privacy policy based on their home-country rather than the users' location. This may happen if EU-based websites comply with the GDPR *tout court*, or if US-based domains fail to comply with the GDPR for the EU visitors, or both. We deal with this issue by differentiating US- and EU-based websites through their domains. In particular, we categorize the ".com" URLs as US-based domains, and ".fr" and ".de" as French and German, respectively, and proceed for the other EU-countries in our dataset analogously. We then estimate Equation (2) separately for US- and EU-based domains. In this way, we check for the presence of spillovers between treatment and control traffic within US- or EU-based domains.

Table 10 collects estimated coefficients from such an experiment. In Columns 1-4 only EU-based domains are used. The estimates are qualitatively the same as in our main analysis. The long-run impact on web traffic is however larger, suggesting (not surprisingly) that EU-based websites might have complied more rigorously with the GDPR. The same holds for user engagement measures, as the long-run effects are generally higher than the ones in Table

7, with the difference being sizeable only for average visit duration. Even when we focus on US-based domains (Columns 5-8) the results are qualitatively stable, implying long-run increasing effects of the GDPR. Estimates are slightly lower in magnitude compared with our main results, but the difference is not particularly large. Hence, also EU traffic for US-based websites have been affected by the GDPR.

Overall, our robustness checks confirm the results in the main analysis, pointing to a reduction in web traffic and user engagement induced by the GDPR, particularly in the long-run. They also validate our empirical design, since we do not detect large biased effects induced by spillovers between treatment and control.

8 Conclusions

In this paper we investigated the effect of the European Union's General Data Protection Regulation on websites' ability to attract visitors and users' interaction with websites. We document an overall average traffic reduction of 15% in the long-run following the GDPR's implementation, and a measurable reduction of user engagement with websites. Traffic from direct visits, organic search, email marketing and social media links, display advertising and referrals dropped significantly after the GDPR. Email marketing and display advertising experienced a large, near 30% reduction of traffic in the long-run.

Our results carry relevance for broader economic policy. They are consistent with the view that users care about privacy and actively opt out of visits as a result of better information about website data handling policies. Paid search traffic – mainly Google search advertisements – was barely affected. Privacy regulations can therefore impact market structure and may increase dependence on large advertising service providers. Our results also indicate that enforcement matters. The effects were amplified considerably following the first large fine issued eight months after full entry into force of the regulation.

We furthermore document an inverted U-shaped relationship between website size and the change in traffic due to privacy regulation. While the smallest and largest websites lost visits, medium websites remained largely unaffected and may have even gained from the GDPR. This confirms pre-GDPR worries about a potential adverse effect of privacy regulation on small firms, but nuance it by finding a negative impact also on large firms.

While there is a positive correlation between traffic and revenues in the internet economy, our results are not directly translatable into welfare effects of the regulation. First, digital platform may better monetize from the remaining users, and this in turn may offset the revenue loss from lower visitors. Second, we do not measure consumer privacy benefits, nor can we reliably convert lost visits into costs. Overall however, it appears to us that additional consumer benefits may easily outweigh the implied losses of website traffic: The GDPR improved privacy online and offline for hundreds of millions of European consumers. Web traffic is probably the domain where consumer reaction to privacy regulation is the largest as some website visits may be deferred. And this loss of visits may not be equivalent to economic harm. For example, the decline of direct website visits may be the result of users' choice to refrain from interacting with privacy-intrusive websites. Since the GDPR enabled such informed choice by mandating that websites obtain informed consent to data policies, even lost website visits may imply net welfare gains.

Overall, our analysis demonstrates that privacy regulation had a measurable impact on the online economy. We take a step towards better understanding how these effects are distributed across different players on various levels of the value chain. Informing policy makers and managers about such effects remains an important task for further research.

References

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

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

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., and Turner, E. (2019). Americans and privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center*, 15:175–190.

BBC (2018). These companies are getting killed by gdpr.

Calzada, J. and Gil, R. (2020). What do news aggregators do? evidence from google news in spain and germany. *Marketing Science*, 39(1):134–167.

Campbell, J., Goldfarb, A., and Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*, 24(1):47–73.

Cherry, M. (2017). Small businesses at risk of being overwhelmed by data protection burden.

CNIL (2019). The CNIL's restricted committee imposes a financial penalty of 50 Million euros against GOOGLE LLC | CNIL.

Coos, A. (2020). Australian government kicks off privacy act review. *Endpoint Protector Blog*.

Dimakopoulos, P. D. and Sudaric, S. (2018). Privacy and platform competition. *International Journal of Industrial Organization*, 61:686–713.

European Commission (2011). Eu: Attitudes on data protection and electronic identity in the european union.

European Commission (2015). Special eurobarometer 431: Data protection.

European Commission (2019). Mythbusting: General data protection regulation - fact sheet.

Gal, M. S. and Aviv, O. (2020). The competitive effects of the gdpr. *Journal of Competition Law & Economics*, 16(3):349–391.

Godinho de Matos, M. and Adjerid, I. (2021). Consumer consent and firm targeting after gdpr: The case of a large telecom provider. *Management Science*.

Goldberg, S., Johnson, G., and Shriver, S. (2019). Regulating Privacy Online: An Economic Evaluation of the GDPR.

Goldfarb, A. and Tucker, C. E. (2011). Privacy regulation and online advertising. *Management science*, 57(1):57–71.

Government of Canada (2020). Fact sheet: Digital charter implementation act, 2020.

Harris, W. (2018). Why marketers need to stop saying “email marketing is dead”.

Hu, X., de Tangil, G. S., and Sastry, N. (2020). Multi-country study of third party trackers from real browser histories. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 70–86. IEEE.

IAPP (2021). The california privacy rights act of 2020. https://iapp.org/resources/article/the-california-privacy-rights-act-of-2020/. (Accessed on 12/12/2021).

Interactive Advertising Bureau (2021). Iab study shows internet economy is transforming the u.s. economy, creating new markets and spurring job growth for large and small businesses.

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*.

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*, 39(1):33–51.

Kottasová, I. (2018). These companies are getting killed by gdpr.

Libert, T., Graves, L., and Nielsen, R. K. (2018). Changes in third-party content on european news websites after gdpr.

Lu, S., Wang, X., and Bendle, N. (2020). Does piracy create online word of mouth? an empirical analysis in the movie industry. *Management Science*, 66(5):2140–2162.

Matte, C., Bielova, N., and Santos, C. (2020). Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe’s transparency and consent framework. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 791–809. IEEE.

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*, pages 1–13.

Ourworldindata.org (2021). Share of the population using the internet. https://ourworldindata.org/grapher/share-of-individuals-using-the-internet?tab=chart. (Accessed on 12/12/2021).

Peukert, C., Bechtold, S., Batikas, M., and Kretschmer, T. (2020). European Privacy Law and Global Markets for Data.

Sabatino, L. and Sapi, G. (2021). Online privacy and market structure: Theory and evidence.

Schmitt, J., Miller, K. M., and Skiera, B. (2021). The impact of privacy laws on online user behavior. *arXiv preprint arXiv:2101.11366*.

Sørensen, J. and Kosta, S. (2019). Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites. pages 1590–1600.

The Economist (2014). Getting to know you.

The Economist (2019). The French fine against Google is the start of a war.

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

Zhuo, R., Huffaker, B., Greenstein, S., et al. (2021). The impact of the general data protection regulation on internet interconnection. *Telecommunications Policy*, 45(2):102083.

9 Tables

Table 1: Summary Statistics

	Mean	**SD**	**Min**	**Max**	**Observations**
Overall					
Traffic	315,112	9,973,771	0	1,970,498,007	4,338,812
Avg. Visit Duration (seconds)	394.45	876.99	0	86,235	3,844,734
Avg. Pages Visited (N)	6.40	9.54	0.10	1,263.59	3,984,617
Bounce Rate	0.45	0.24	0.00	1.00	3,839,368
Traffic by channel					
Direct	1,485,277	26,116,612	17	1,970,498,007	619,999
Display Advertising	13,611.48	92,249.68	0	6,042,370	619,760
Email	55,130.18	614,207.90	0	83,178,035	619,959
Organic Search	443,680.79	3,275,475	0	225,132,942	619,998
Paid Search	21,493.86	191,691.46	0	21,070,592	619,106
Referrals	103,117.49	618,041.18	0	28,195,818	619,995
Social	82,916.88	896,998.53	0	51,305,963	619,995
Traffice by size					
Small	46,461.67	99,237.95	0	5,154,861	1,956,409
Medium	146,163.58	330,465.12	0	9,629,733	1,241,459
Large	317,105.24	1,457,990	0	40,895,073	986,016
Giant	5,048,721	52,419,507	0	1,970,498,007	154,928

This table shows summary statistics for the main variables used in the empirical analysis. Source: authors' calculations from Similarweb data.

Table 2: Parallel Trends Estimates

	(1)	(2)	(3)	(4)
VARIABLES	Traffic	Avg Visit Duration	Avg Pages Visited	Bounce Rate
Trend$_{EU}$	-0.0004	-0.0016***	-0.0011***	0.0005***
	(0.0006)	(0.0003)	(0.0002)	(0.0001)
Trend$_{US}$	-0.0020***	-0.0010**	-0.0011***	0.0005***
	(0.0007)	(0.0005)	(0.0002)	(0.0001)
H_0: Trend$_{EU}$ =Trend$_{US}$				
F-test	4.40	0.88	0.05	0.09
Observations	906,899	807,279	834,344	806,740
R-squared	0.719	0.422	0.601	0.497

Presented are OLS estimated coefficients of Equation (5). The dependent variable is the $\ln(x+1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. The Post dummy takes value 1 from GDPR introduction onward, while EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Diff-in-Diff on Web Traffic

VARIABLES	(1) Traffic	(2) Traffic	(3) Traffic	(4) Traffic	(5) Traffic	(6) Traffic	(7) Traffic
Post × EU	-0.104***	-0.104***	-0.104***	-0.104***	-0.104***	-0.106***	-0.064***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
Post	0.034***						
	(0.011)						
Europe	-2.190***	-2.190***					
	(0.044)	(0.044)					
Domain FE	X	X	X	X			
Channel FE	X	X	X				
Time FE		X	X				
Country FE			X				
Channel-Time FE				X	X	X	X
Country-Channel FE				X	X		
Domain-Channel FE					X		
Domain-Channel-Country FE						X	X
Domain-Time FE							X
Observations	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812
R-squared	0.694	0.694	0.700	0.706	0.928	0.942	0.956

Presented are OLS estimated coefficients of Equation (1). The dependent variable is the $\ln(x+1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. The Post dummy takes value 1 from GDPR introduction onward, while EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Diff-in-Diff on Web Traffic in the Short and Long Run

VARIABLES	(1) Traffic	(2) Traffic	(3) Traffic	(4) Traffic	(5) Traffic	(6) Traffic	(7) Traffic
Post^S × EU	-0.040***	-0.040***	-0.040***	-0.040***	-0.040***	-0.042***	-0.027**
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)
Post^L × EU	-0.157***	-0.157***	-0.157***	-0.157***	-0.157***	-0.159***	-0.094***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Post^S	0.026***						
	(0.010)						
Post^L	0.042***						
	(0.014)						
EU	-2.190***	-2.190***					
	(0.044)	(0.044)					
Domain FE	X	X	X	X			
Channel FE	X	X	X				
Time FE		X	X				
Country FE			X				
Channel-Time FE				X	X	X	X
Country-Channel FE				X	X		
Domain-Channel FE					X		
Domain-Channel-Country FE						X	X
Domain-Time FE							X
Observations	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812
R-squared	0.694	0.694	0.700	0.706	0.928	0.942	0.956

Presented are OLS estimated coefficients of Equation (2). The dependent variable is the $\ln(x+1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Heterogeneous Effects on Web Traffic by Traffic Channel

VARIABLES	(1) Traffic	(2) Traffic	(3) Traffic	(4) Traffic	(5) Traffic	(6) Traffic
Post^S × EU × Direct	-0.039***	-0.039***	-0.038***	-0.040***	-0.041***	-0.026**
	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)	(0.013)
Post^S × EU × Display Advertising	-0.053**	-0.053**	-0.052**	-0.055**	-0.058**	-0.043*
	(0.026)	(0.026)	(0.026)	(0.025)	(0.025)	(0.024)
Post^S × EU × Email	-0.079***	-0.079***	-0.079***	-0.075***	-0.077***	-0.062***
	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)	(0.017)
Post^S × EU × Organic Search	-0.010	-0.010	-0.010	-0.007	-0.009	0.006
	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)	(0.013)
Post^S × EU × Paid Search	-0.029	-0.028	-0.028	-0.038	-0.040	-0.026
	(0.030)	(0.030)	(0.030)	(0.030)	(0.030)	(0.029)
Post^S × EU × Referrals	-0.028*	-0.028*	-0.028*	-0.028*	-0.029*	-0.014
	(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.017)
Post^S × EU × Social	-0.042***	-0.042***	-0.042***	-0.040***	-0.041***	-0.026*
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)
Post^L × EU × Direct	-0.088***	-0.088***	-0.088***	-0.088***	-0.089***	-0.023
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.019)
Post^L × EU × Display Advertising	-0.381***	-0.381***	-0.380***	-0.381***	-0.386***	-0.320***
	(0.037)	(0.037)	(0.037)	(0.036)	(0.036)	(0.034)
Post^L × EU × Email	-0.291***	-0.291***	-0.291***	-0.287***	-0.288***	-0.223***
	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.025)
Post^L × EU × Organic Search	-0.073***	-0.073***	-0.073***	-0.071***	-0.072***	-0.007
	(0.016)	(0.016)	(0.016)	(0.015)	(0.015)	(0.019)
Post^L × EU × Paid Search	-0.068*	-0.068*	-0.067*	-0.075*	-0.079**	-0.013
	(0.040)	(0.040)	(0.040)	(0.039)	(0.039)	(0.039)
Post^L × EU × Referrals	-0.079***	-0.079***	-0.079***	-0.080***	-0.081***	-0.016
	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.024)
Post^L × EU × Social	-0.123***	-0.123***	-0.123***	-0.120***	-0.121***	-0.055***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.020)
Domain FE	X	X	X			
Channel FE	X					
Channel-Time FE		X	X	X	X	X
Country-Channel FE			X	X		
Domain-Channel FE				X		
Domain-Channel-Country FE					X	X
Domain-Time FE						X
Observations	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812	4,338,812
R-squared	0.696	0.696	0.707	0.928	0.942	0.956

Presented are OLS estimated coefficients of Equation (2) interacted with channel-specific dummies. Channels are categorized in Section 4, and identify the way through which the website is reached. The dependent variable is the $\ln(x+1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 6: Diff-in-Diff on Web Traffic by Website Size

	Small		Medium		Large		Giants	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Traffic	Traffic	Traffic	Traffic	Traffic	Traffic	Traffic	Traffic
Post^S × EU	-0.142***	-0.143***	0.024	0.024	-0.008	-0.009	-0.046	-0.046
	(0.025)	(0.025)	(0.016)	(0.016)	(0.016)	(0.016)	(0.031)	(0.031)
Post^L × EU	-0.394***	-0.394***	-0.038	-0.038	-0.073***	-0.073***	-0.155***	-0.155***
	(0.035)	(0.035)	(0.024)	(0.024)	(0.024)	(0.024)	(0.047)	(0.047)
Domain FE	X		X		X		X	
Channel FE	X		X		X		X	
Time FE	X		X		X		X	
Country FE	X		X		X		X	
Channel-Time FE		X		X		X		X
Country-Channel FE		X		X		X		X
Domain-Channel FE		X		X		X		X
Observations	1,956,409	1,956,409	1,241,459	1,241,459	986,016	986,016	154,928	154,928
R-Squared	0.658	0.925	0.704	0.936	0.725	0.917	0.819	0.909

Presented are OLS estimated coefficients of Equation (2). The dependent variable is the $\ln(x+1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Diff-in-Diff on User Engagement

	(1)	(2)	(3)
VARIABLES	Avg Visit Duration	Avg Pages Visited	Bounce Rate
Post^S × EU	-0.017***	-0.007**	0.000
	(0.006)	(0.003)	(0.001)
Post^L × EU	-0.035***	-0.015***	0.006***
	(0.007)	(0.005)	(0.002)
Channel-Time FE	X	X	X
Country-Channel FE	X	X	X
Domain-Channel FE	X	X	X
Observations	3,844,720	3,984,614	3,839,355
R-squared	0.550	0.721	0.697

Presented are OLS estimated coefficients of Equation (2). In the first two columns, the dependent variable is the $\ln(x+1)$, where x is the average visit duration measured in seconds (Column 1) and the number of pages visited (Column 2), while in Column 3 the dependent variable is the bounce rate for domain i, for traffic coming from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Heterogeneous Effects on User Engagement by Traffic Channel

	(1)	(2)	(3)
VARIABLES	Avg Visit Duration	Avg Pages Visited	Bounce Rate
Post^S × EU × Direct	-0.003	-0.002	-0.002*
	(0.005)	(0.004)	(0.001)
Post^S × EU × Display Advertising	0.009	-0.001	-0.002
	(0.022)	(0.009)	(0.004)
Post^S × EU × Email	-0.051***	-0.026***	0.009***
	(0.012)	(0.006)	(0.002)
Post^S × EU × Organic Search	-0.010*	-0.010**	0.003
	(0.006)	(0.004)	(0.002)
Post^S × EU × Paid Search	-0.042*	0.003	-0.004
	(0.022)	(0.010)	(0.004)
Post^S × EU × Referrals	-0.001	0.005	-0.006***
	(0.008)	(0.005)	(0.002)
Post^S × EU × Social	-0.025***	-0.013**	0.002
	(0.010)	(0.005)	(0.002)
Post^L × EU × Direct	-0.002	0.002	0.001
	(0.007)	(0.006)	(0.002)
Post^L × EU × Display Advertising	0.010	0.000	0.003
	(0.024)	(0.010)	(0.004)
Post^L × EU × Email	-0.090***	-0.057***	0.022***
	(0.015)	(0.008)	(0.003)
Post^L × EU × Organic Search	-0.013*	-0.008	0.005**
	(0.007)	(0.005)	(0.002)
Post^L × EU × Paid Search	-0.079***	-0.019*	-0.001
	(0.023)	(0.011)	(0.005)
Post^L × EU × Referrals	-0.028***	-0.004	-0.000
	(0.010)	(0.007)	(0.003)
Post^L × EU × Social	-0.052***	-0.020***	0.007***
	(0.012)	(0.007)	(0.002)
Channel-Time FE	X	X	X
Country-Channel FE	X	X	X
Domain-Channel FE	X	X	X
Observations	3,844,720	3,984,614	3,839,355
R-squared	0.550	0.721	0.697

Presented are OLS estimated coefficients of Equation (2) interacted with channel-specific dummies. Channels are categorized in Section 4, and identify the way through which the website is reached. In the first two columns, the dependent variable is the $\ln(x+1)$, where x is the average visit duration measured in seconds (Column 1) and the number of pages visited (Column 2), while in Column 3 the dependent variable is the bounce rate for domain i, for traffic coming from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

Table 9: Robustness Check - No Multinational Traffic

	(1)	(2)	(3)	(4)
VARIABLES	Traffic	Avg Visit Duration	Avg Pages Visited	Bounce Rate
Post^S × EU	-0.037**	-0.032***	-0.015***	0.004**
	(0.017)	(0.009)	(0.005)	(0.002)
Post^L × EU	-0.179***	-0.053***	-0.022***	0.010***
	(0.024)	(0.011)	(0.007)	(0.002)
Channel-Time FE	X	X	X	X
Country-Channel FE	X	X	X	X
Domain-Channel FE	X	X	X	X
Observations	1,864,121	1,658,465	1,710,556	1,649,193
R-squared	0.938	0.574	0.739	0.731

Presented are OLS estimated coefficients of Equation (2). The dependent variable is the $\ln(x+1)$, where x is the number of visits (Column 1), the average visit duration measured in seconds (Column 2), and the number of pages visited (Column 3), while in Column 3, the dependent variable is the bounce rate for domain i, for traffic coming from channel k and country c, at time (weekly date) t of domain i from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

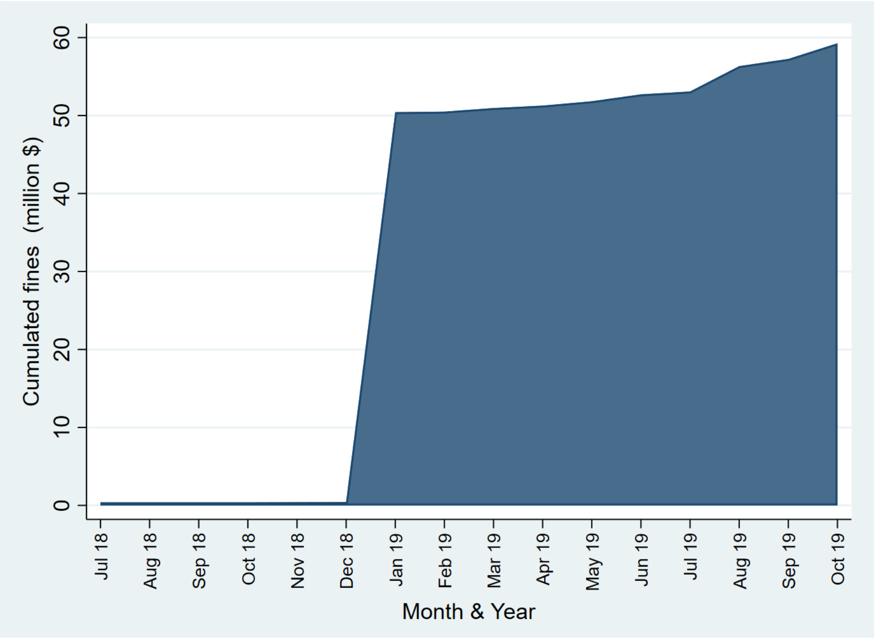
Table 10: Robustness Check - Assessing Potential Spillovers

	EU Domains Only				**US Domains Only**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Traffic	Avg Visit Duration	Avg Pages Visited	Bounce Rate	Traffic	Avg Visit Duration	Avg Pages Visited	Bounce Rate
$\text{Post}^S \times \text{EU}$	-0.040*	-0.015	-0.012**	0.002	-0.036***	-0.013	-0.003	-0.002
	(0.024)	(0.010)	(0.006)	(0.002)	(0.014)	(0.009)	(0.005)	(0.002)
$\text{Post}^L \times \text{EU}$	-0.246***	-0.053***	-0.017**	0.006**	-0.128***	-0.025**	-0.015**	0.006**
	(0.032)	(0.013)	(0.007)	(0.003)	(0.021)	(0.011)	(0.007)	(0.002)
Channel-Time FE	X	X	X	X	X	X	X	X
Country-Channel FE	X	X	X	X	X	X	X	X
Domain-Channel FE	X	X	X	X	X	X	X	X
Observations	2,026,044	1,760,007	1,830,579	1,758,138	2,312,768	2,084,713	2,154,035	2,081,217
R-squared	0.925	0.528	0.703	0.693	0.929	0.569	0.736	0.701

Presented are OLS estimated coefficients of Equation (2). The dependent variable is the $\ln(x+1)$, where x is the number of visits (Columns 1 and 5), the average visit duration measured in seconds (Columns 2 and 6), and the number of pages visited (Columns 3 and 7), while in Columns 4 and 8, the dependent variable is the bounce rate for domain i, for traffic coming from channel k and country c, at time (weekly date) t of domain i from channel k and country c, at time (weekly date) t. Post^S is a dummy taking value 1 from GDPR introduction to the week before the first large fine imposed by the French privacy authority to Google (21st of January 2109), while Post^L activates from that date onward. EU is an indicator that identifies traffic generated from European countries. Robust standard errors clustered by domain in parenthesis. *** p<0.01, ** p<0.05, * p<0.1

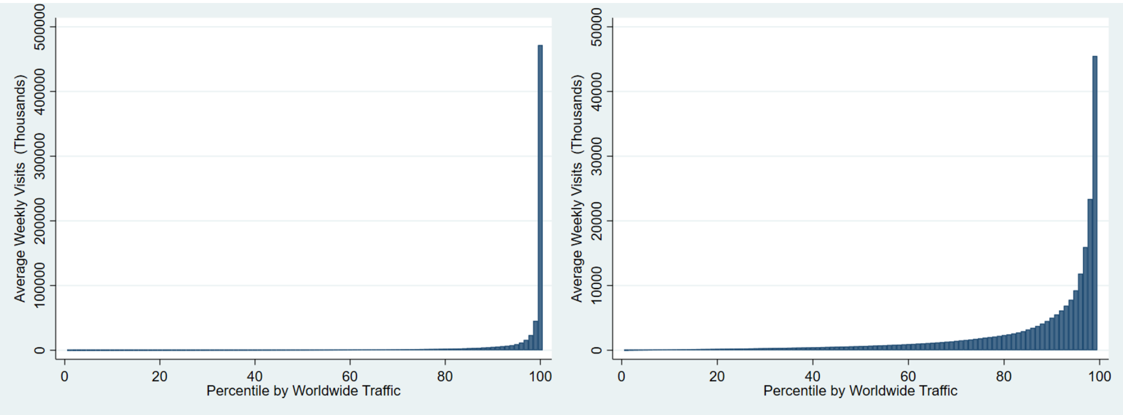
10 Figures

Figure 1: Cumulated Fines under the GDPR



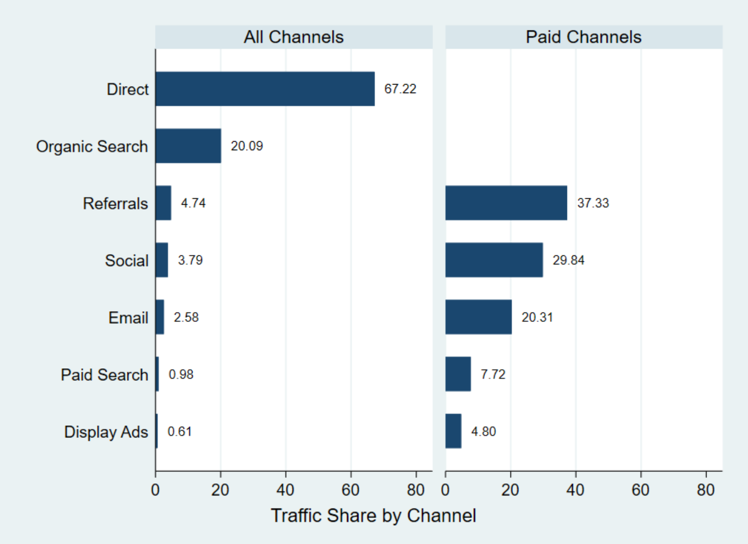
This figure shows the cumulated fines following GDPR infringements. The jump in January 2019 refers to the financial penalty imposed by the French privacy authority (CNIL) against Google. Source: GDPR Enforcement Tracker at https://www.enforcementtracker.com/

Figure 2: Concentration in the Digital Market



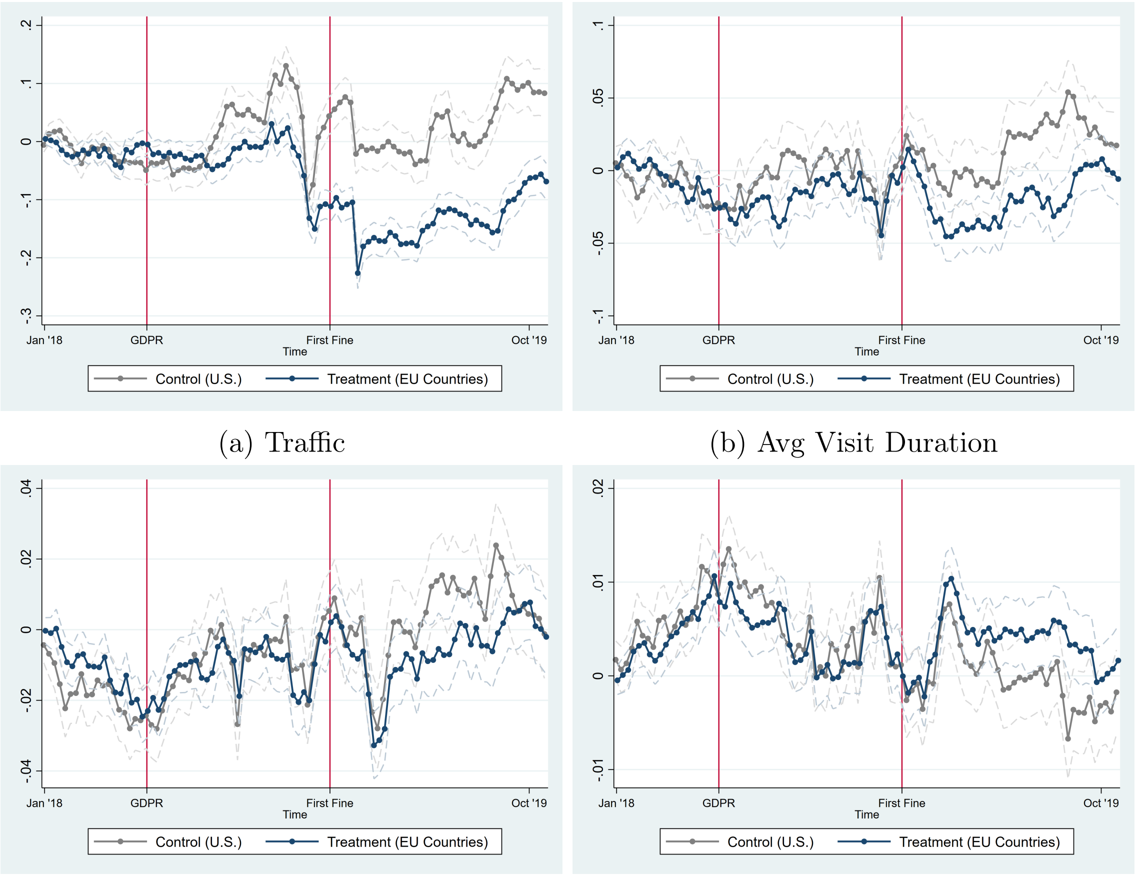
This figure shows average visits (in thousands) for each percentile of the distribution of worldwide traffic before GDPR. On the left, all percentiles are listed. On the right panel, the top 1st percentile is excluded. Source: Similarweb data.

Figure 3: Distribution of Web Traffic across Channels



This figure shows the traffic share by channel. On the left, all traffic is considered when determining each channel's share. On the right, only paid traffic is considered. Source: Similarweb data.

Figure 4: Parallel Trends: Graphical Evidence



(a) Traffic

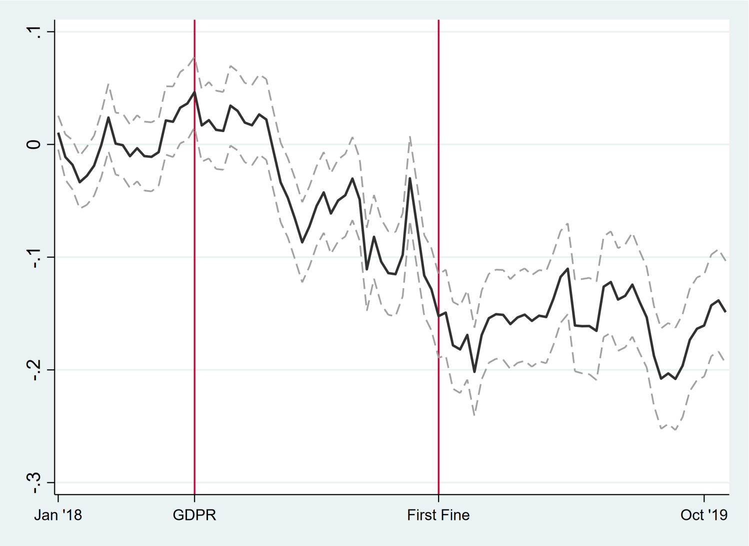
(b) Avg Visit Duration

(c) Avg Pages Visited

(d) Bounge Rate

This figure shows OLS coefficients associated to the time dummies τ_t for both the US (gray line) and Europe (blue line) of Equation (4), together with the corresponding 95% confidence interval. The first-period dummy is excluded to avoid multicollinearity. Robust standard errors are clustered by domain.

Figure 5: Web Traffic: Dynamic Effect



This figure shows the $\hat{\beta}_{1,t}$ coefficients from Equation (3) with the associated 95% confidence interval. The first-period dummy is excluded to avoid multicollinearity. The dependent variable is the $\ln(x + 1)$, where x is the number of visits of domain i from channel k and country c, at time (weekly date) t. Robust standard errors are clustered by domain.