REGULATING PRIVACY ONLINE: AN ECONOMIC EVALUATION OF THE GDPR*

Samuel G. Goldberg, Garrett A. Johnson and Scott K. Shriver

June 23, 2021

Abstract

Modern websites rely on personal data to design better content and to market themselves. The European Union's General Data Protection Regulation (GDPR) was intended to make access to such personal data much more difficult, with the goal of protecting user privacy. We examine the GDPR's impact on website pageviews and revenue for 1,084 diverse online firms using data from Adobe's website analytics platform. Using a panel differences approach, we find a reduction of approximately 12% in both EU user website pageviews and website e-commerce revenue recorded by the platform after the GDPR's enforcement deadline. To understand the policy's impact, we must separate the GDPR's effects on real outcomes from its impact on data recording. We derive informative bounds on both components by first examining site-usage patterns of selected users that we continue to observe post-GDPR. We bound changes to data recording, attributed to users opting-out of data collection, to be between 4% and 15%. This implies that at least 7% of our GDPR estimates (0.8% of total) arise from the consent effect on data recording. We then find larger effects on traffic from email and display ad marketing channels—specific targets of the GDPR. We conclude that at least 8% of our GDPR estimates (0.4% of total) arise from the GDPR's real effect on marketing alone. However, we do not find evidence that consent interfaces dissuade users from browsing sites.

JEL Codes: L51, L86, M31, M38

*Corresponding Author: Samuel Goldberg, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston IL 60208. <Samuel.Goldberg@kellogg.northwestern.edu> Ph: (610) 291-8450; Johnson: Questrom School of Business, Boston University <garjoh@bu.edu>; Shriver: Leeds School of Business, University of Colorado – Boulder <Scott.Shriver@colorado.edu>. The authors would like to thank Adobe for providing the data, Taylor Schreiner for his support within Adobe, Anita Rao, and Catherine Tucker for helpful discussions, Dirk Bergemann, Andrey Fradkin, Brett Gordon, Laura Kornish, Tesary Lin and Brad Shapiro for comments, as well as seminar participants at Boston University, Harvard Business School, INSEAD, Kellogg School of Management, Lehigh University, Marketing Science, and NBER Digitization Summer Institute 2019. The research was conducted while Samuel Goldberg was an employee of Adobe and partially funded by MSI research grant #4000783. All errors are our own.

1 Introduction

Personal data is a fundamental input of the modern digital economy. Firms leverage personal data for a variety of activities, including product recommendations, advertising and pricing. Concerns over invasive and opaque data collection practices have fueled a political debate about the consumer's right to privacy and how best to regulate it. The European Union's (EU) General Data Protection Regulation (GDPR) seeks to protect individual online privacy by reigning in firms' use of personal data. The GDPR is one of the most significant privacy regulations in effect today and has served as a model for subsequent privacy regulations such as California's Consumer Privacy Act. While privacy may be valuable to individuals, by limiting the use of personal data the GDPR may have the unintended consequence of harming the online firms that represent 4% of the EU's economy and 10% of its retail sales.[1] If so, policy-makers must weigh the economic cost of regulation against its privacy benefit. This paper presents a broad-based, empirical study of the GDPR's economic impact on online firms. We examine the effect of the GDPR on website traffic and on e-commerce revenue, as advertising and e-commerce sales are the main revenue sources for online firms.

A key contribution of this paper is to distinguish between the GDPR's impact on real economic activity and the process of recording economic outcomes. Extant policy research examines how regulation affects economic outcomes, but seldom contends with policies that also affect how those outcomes are recorded. Privacy policy creates missingness by construction: individuals can forbid their data from being recorded. Under the GDPR, recorded economic outcomes may *appear to fall* because some individuals do not consent to data sharing. At the same time, the GDPR could affect *real economic outcomes*, for instance, because the regulation restricts personalized marketing. Privacy regulation thus creates an inference problem: data protection can obscure both how and how much the regulation affects the economy, which complicates policy evaluation. Policymakers may view mere reductions in data recording as evidence that privacy regulation is successful. However, reductions in recorded outcomes could hide true underlying harm to the economy from policymakers. We construct informative bounds on the real and data recording effects of the GDPR in the spirit of Manski (2007), as has been applied in some other policy domains (e.g. Manski (2014); Manski and Pepper (2018)).

We show that the GDPR reduces *recorded* pageviews and e-commerce revenue by 12%. We note larger drops in users arriving from personalized marketing channels that are most directly impacted by the GDPR. Under weak assumptions, we show the real effect of the GDPR reduced *real* website pageviews by 0.4% and *real* e-commerce revenue by 0.6%. We

[1] https://www.retailresearch.org/online-retail.html

stress that these are conservative bounds. The alternative explanation for the observed reduction in recorded outcomes is the GDPR's consent effect on data recording. We estimate a lower bound for this consent effect of 0.8%: meaning that our data are also consistent with a larger GDPR effect on real outcomes. Our economic harm estimate contributes to the literature on the economic cost of regulation (Joskow and Rose, 1989; Laffont, 1994). We note that individual firms likely can not distinguish a small GDPR effect from idiosyncratic shocks and policymakers also lack the granular, online-specific data necessary to detect a small (real) effect of the GDPR. Moreover, our findings comport with extant research on the GDPR that generally finds small effects on websites (e.g. Aridor et al. 2020; Johnson et al. 2020). These small effects should to be interpreted in the broader context of both low regulatory compliance and low enforcement. EU regulator sweeps have found that EU websites almost universally do not comply with the GDPR's strict opt-in requirement (Autoriteit Persoonsgegevens, 2019; Data Protection Commission, 2020b). In contrast, sites that required GDPR-compliant consent found that only about 10% of users provided opt-in consent (Snelders et al., 2020) so that recorded site analytics data fell over 90% (Information Commissioner's Office, 2019).

We leverage proprietary data from Adobe Analytics to understand the implications of the GDPR for the online economy. Governments are turning to private data providers as government data can lack sufficient granularity to investigate certain policy questions. For instance, the United States Bureau of Labor Statistics, Federal Reserve, and Census Bureau all partner with Adobe Analytics to obtain real-time economic data (Adobe, 2020). These data are used to study important policy issues like inflation (Goolsbee and Klenow, 2018). Our data consists of recorded outcomes from over 1,000 websites from such diverse industries as media, travel, beauty, health, and retail. We observe 77 of the top 1,000 websites globally as well as over 700 sites below the top 100,000. The data separates the web traffic by EU residents who are protected by the GDPR. Our analysis focuses on two key metrics of the online economy: pageviews and e-commerce revenue.[2] We observe over 4.4 billion weekly pageviews and about $0.75 billion in e-commerce revenue, per week, from EU users alone—representing 12% of total European e-commerce. For advertising-supported websites, pageviews indicate a site's supply of ads and ad revenue. E-commerce revenue contributes to overall economic activity as 17% of EU enterprises sold online in 2018 (EuroStat, 2020). The data also identify the channels (e.g. search engine, email, display ad) through which users arrive at the website, which we use to examine how the GDPR affected online marketing to EU users.

[2]Our revenue measures consists of all revenue collected by the site from selling goods online, but excludes advertising revenue.

Like past economic studies of privacy regulation (e.g. Goldfarb and Tucker, 2011; Miller and Tucker, 2009), we use the timing of regulatory enforcement as an event study. The GDPR affects firms at the same time, so selecting a control group is challenging. Our preferred control group is the same set of sites in the previous year. This control group then captures the seasonal pattern of EU user traffic, which is specific to these firms. Our primary analyses use a panel differences (similar in spirit to a differences-in-differences model) approach to identify the impact of the GDPR on recorded pageviews and recorded e-commerce revenue.[3] We estimate that recorded pageviews fall 11.7% post-GDPR, on average across all sites, or 15,043 pageviews per week for the median site. Among e-commerce sites, recorded revenue falls 13.3%, or $9,227 per week for the median site. These results are robust to a variety of specifications including using North American user outcomes as a control. While the above results suggest that the GDPR has changed online outcomes, they do not disentangle the role of data recording from real economic harms.

The GDPR can impact recorded web outcomes through several mechanisms. European privacy regulators state that sites must obtain *user consent* for collecting site analytics data (Article 29 Data Protection Working Party, 2012; Information Commissioners Office, 2019; Data Protection Commission, 2020a). The most adopted consent management platform indicates that over 90% of users provide consent (Quantcast, 2018). If up to 10% of users do not provide consent, this could explain some of the 11.7% reduction in recorded pageviews. Furthermore, privacy regulation may limit firm's *marketing* ability. Privacy regulation has been documented to negatively impact marketing effectiveness, both in the context of the EU's 2003 ePrivacy Directive (Goldfarb and Tucker, 2011) and the GDPR (de Matos and Adjerid, 2019; Aridor et al., 2020). Marketing may be less effective particularly on channels like email and display advertising that rely on personal data. This marketing mechanism would lead to a decrease in real—and not merely recorded—site outcomes. Consent dialogs on websites may also reduce real site outcomes by degrading the user browsing experience. In particular, GDPR consent dialogs may create *privacy frictions* for users that discourage users from browsing the site further. Finally, though some firms may stop sharing data with Adobe to conform with the GDPR's principle of data minimization, we rule this mechanism out by excluding such firms from our analysis.

We quantify the relative contribution of various mechanisms. We find that a combination of consent and marketing drive the reduction in recorded site outcomes. We begin with the role of consent, which limits the share of data recorded by Adobe. We use a simple selection

[3]Formally, differences-in-differences requires contemporaneous observation of cross-sectional units that are not exposed to the treatment of interest. The GDPR's scope and coordinated rollout severely limits the construction of a true control group, since websites serving EU citizens (virtually all major sites) are subject to the GDPR.

argument to construct bounds on the role of consent after the GDPR. We estimate non-consent rates are between 4.1% and 15.4% of site visits in the full sample and between 5.3% and 17.7% for the e-commerce sample, which generally align with the Quantcast (2018) consent figures. In particular, consent explains a minimum of 6.9% of our GDPR estimate for recorded pageviews representing 0.8% of total pageviews.

The remainder of the GDPR estimate could be due to a real effect on EU web traffic, though conservative estimates suggest a small but non-zero effect. First, we use data on the user's arrival channel to look for evidence that marketing has been impacted by the GDPR. We find that the GDPR had a greater impact on users arriving through data-intensive marketing channels like display and email advertising. Net of the consent effect, we calculate that marketing mechanisms reduce real pageviews by approximately 0.4% per week. For revenue, marketing mechanisms reduce real revenue by approximately 0.6% per week. Thus, the real effect of the GDPR through the display and email channels alone represents 9.4% of the estimated GDPR effect on pageviews and 7.6% of the estimated GDPR effect on e-commerce revenue . Second, the privacy frictions mechanism could also affect real outcomes. We argue privacy frictions would increase the share of site visits where the user browses a single page then leaves, a behavior known as bouncing. Since bounce rates do not materially change post-GDPR, we conclude that the privacy frictions mechanism does not play a substantial role in post-GDPR outcomes.

We show the GDPR has differential effects on sites by size and regulatory strictness. Extant research highlights anti-competitive effects of the GDPR (Johnson et al., 2020; Peukert et al., 2020). We find evidence that the GDPR has a greater effect on smaller e-commerce sites, though we see no difference by size in the full sample. In particular, smaller e-commerce sites see twice the decline in recorded revenue (-17.4%) than larger sites (-8.9%). We show this decline arises from smaller e-commerce sites having a harder time obtaining consent, which validates a key theoretical prediction in Campbell et al. (2015). We also consider the moderating effect of regulatory strictness. The GDPR harmonized regulation within the EU, but enforcement is partly at the country level where regulator resources vary. We show that site beliefs about data protection regulatory strictness—as proxied by a EU survey of firms—moderate the effect of the GDPR. Specifically, a one standard deviation increase in the regulatory strictness index corresponds to 2.1% lower recorded pageviews and 4.5% lower recorded revenue. This finding aligns with other research showing a correlation between regulatory strictness and the economic effects of the GDPR (Jia et al., 2018; Johnson et al., 2020).

Our work expands the literature on the economic implications of privacy regulation. Pre-GDPR studies of privacy regulation show that sectoral privacy laws can slow technol-

ogy diffusion (Miller and Tucker, 2009; Adjerid et al., 2016) and the EU's 2009 e-Privacy Directive reduced advertising effectiveness (Goldfarb and Tucker, 2011). Emergent research on the GDPR reveals further economic consequences of privacy regulation. Jia et al. (2018, 2019) note that venture capital investment falls in the EU post-GDPR. Theoretical research suggests that the GDPR's privacy rights can strengthen consumers at the expense of firms (Ke and Sudhir, 2020). Computer science research shows that websites use fewer third-party vendors—which often rely on personal information—after the GDPR (e.g. Libert et al. 2018; Sørensen and Kosta 2019), though Zhuo et al. (2019) show no effect of the GDPR at the Internet infrastructure level. Johnson et al. (2020) and Peukert et al. (2020) show that this reduction in third-party vendors favored large vendors and thereby increased market concentration.

Aridor et al. (2020) and Lefrere et al. (2020) are closest to our work in that they also examine how the GDPR affects websites. Lefrere et al. (2020) examine 5,000 web publishers and find that traffic falls about 4% on EU sites post-GDPR relative to US sites. Aridor et al. (2020) examine traffic on travel sites as recorded by a third-party intermediary that sells travel advertising. Their findings complement ours in that both papers estimate similar-sized effects on recorded site traffic and find that recorded outcomes are favorably selected post-GDPR. Aridor et al. (2020) attribute the GDPR effect solely to consent, whereas we explore alternative mechanisms by leveraging our greater cross-section of sites as well as data on visits by marketing channel. Aridor et al. (2020) instead leverage their ad revenue and user scoring data to show that, though the GDPR hinders firm's marketing (see also de Matos and Adjerid, 2019), this is somewhat offset by the favorable selection of consenting users.

This paper is also related to a growing literature on the use of analytics and internet data to measure the economy. Early studies focused on the value of analytics and data-driven decision-making to firms (Bresnahan et al., 2002; Brynjolfsson et al., 2011). Recent work by Berman and Israeli (2020) shows positive returns to firms that adopt a data analytics platform. A handful of studies use web analytics data—including price and transaction data—to measure the online economy. Several researchers estimate the value of the internet and e-commerce (Cavallo and Rigobon, 2016; Dolfen et al., 2019; Brynjolfsson and Oh, 2012). Budak et al. (2016) is closest to our study as they leverage browsing history to assess how e-commerce sites depend on online advertising to generate traffic. We contribute to this literature by tackling missingness in web analytics data due to user consent.

The paper proceeds as follows. Section 2 provides an overview of the GDPR. In Section 3, we explain how the GDPR impacts recorded outcomes and describe our data. In Section 4, we explain the empirical methods we use to measure GDPR-related effects. We present our

main empirical estimates as well as a variety of robustness checks in Section 5. Section 6 explains and quantifies the contribution of various GDPR mechanisms to real and recorded economic outcomes. Section 7 concludes.

2 GDPR background

The European Union (EU) passed the General Data Protection Regulation (GDPR) in April of 2016 and enforcement began on May 25, 2018, giving firms two years to prepare. The GDPR protects the collection, processing, and use of personal information of EU residents. The GDPR expands the definition of personal information to include individual-specific data like cookie identifiers and IP addresses. Fines in the case of non-compliance can reach the larger of 20 million euros or 4% of global revenue. The GDPR regulators split enforcement between country-specific regulators and an EU-level regulator.

Data minimization is a guiding principle of the GDPR: firms must limit the personal data they collect and use. The GDPR accords data rights to EU residents including the right to access, correct, erase, and port their personal data. Firms that process personal data must invest in systems and processes to fulfill these rights-based responsibilities. These firms are also required to audit internal data processes and to appoint a data protection officer to oversee compliance activities. Firms must encrypt and anonymize personal data as well as promptly notify both the regulator and affected individuals after a data breach. These obligations impose significant compliance costs on firms. Many firms spent over 10 million dollars annually to comply with the law and many came into compliance after May 25, 2018 (PricewaterhouseCoopers, 2018). The GDPR requirements incentivize firms to minimize their data processing, including for activities like marketing.

The GPDR lays out legal bases under which a firm may process personal data. EU regulators clarified that individual consent is the most appropriate basis for websites and their vendors to process user data — including for web analytics purposes (Article 29 Data Protection Working Party, 2012; Information Commissioners Office, 2019; Data Protection Commission, 2020a).[4] Under the GDPR, valid consent must be affirmative (no pre-checked boxes), freely given, granular to the purpose of processing (e.g. website analytics, behavioral advertising), and must list all third parties who process the data (European Data Protection

[4]European regulators have long signaled some openness to web analytics that exclusively use the site's data (i.e. using first party cookies), but concern about web analytics vendors that combine user data across site (i.e. using third party cookies) (Article 29 Data Protection Working Party, 2012). In 2020, the French regulator opened the door for sites to collect web analytics data without opt-in consent under the French Data Protection Act provided that this data is not combined with off-site data (Commission Nationale de l'Informatique et des Libertés, 2020b).

Board, 2020). In practice, many websites use choice architecture ("agree to continue") to maximize consent rates and some rely on the legal basis of "legitimate interest."[5] EU regulators object to these practices, but have delayed enforcement actions until 2020 or 2021 (Commission Nationale de l'Informatique et des Libertés, 2020a; Data Protection Commission, 2020b).

3 Data

Our study leverages proprietary data from the Adobe Analytics platform to study the GDPR. Section 3.1 overviews the services and data that web analytics vendors provide. Section 3.2 explains how GDPR affects the data that web analytics platforms record. Section 3.3 describes our panel of online firms and how we construct the panel. Finally, Section 3.4 provides summary statistics for our key outcomes.

3.1 Web analytics overview

Online firms use web analytics tools to understand the characteristics and behaviors of their site visitors. Our data is provided by Adobe Analytics, a leading web analytics vendor (Forrester, 2017). Web analytics vendors like Adobe Analytics provide technology for websites to track users who browse their sites. In order to implement Adobe Analytics, the firm adds code to their website which sends a data-rich ping to Adobe's servers whenever a user visits. These pings contain a unique user ID, website ID, webpage information, and generate a timestamp[6]. Adobe then aggregates the ping data into an *analytics dashboard*. Analytics dashboards are the primary unit of analysis in this paper. Dashboards reveal traffic and revenue performance aggregated over time and broken down, for instance, to the user's country of residence.[7] Online firms find these dashboards to be a source of customer insight that can increase revenue (Berman and Israeli, 2020). For instance, changes in web analytics performance may alert the firm to opportunities to improve its website design or marketing activities. Even the French privacy regulator views site analytics data as "in many cases essential for the proper functioning of the site" (Commission Nationale de l'Informatique et

[5]Legitimate interest may be claimed when "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject" (GDPR (6)(1)(f)). "Legitimate interest" is considered a risky compliance strategy for firms given the need to establish a balance of interests among stakeholders.

[6]We do not observe the unique user ID in our data, but simply aggregated site-level statistics. Thus, we can not track individual user behavior over time.

[7]Refer to the Adobe Analytics (Adobe, 2019) documentation for technical information.

des Libertés, 2020b).

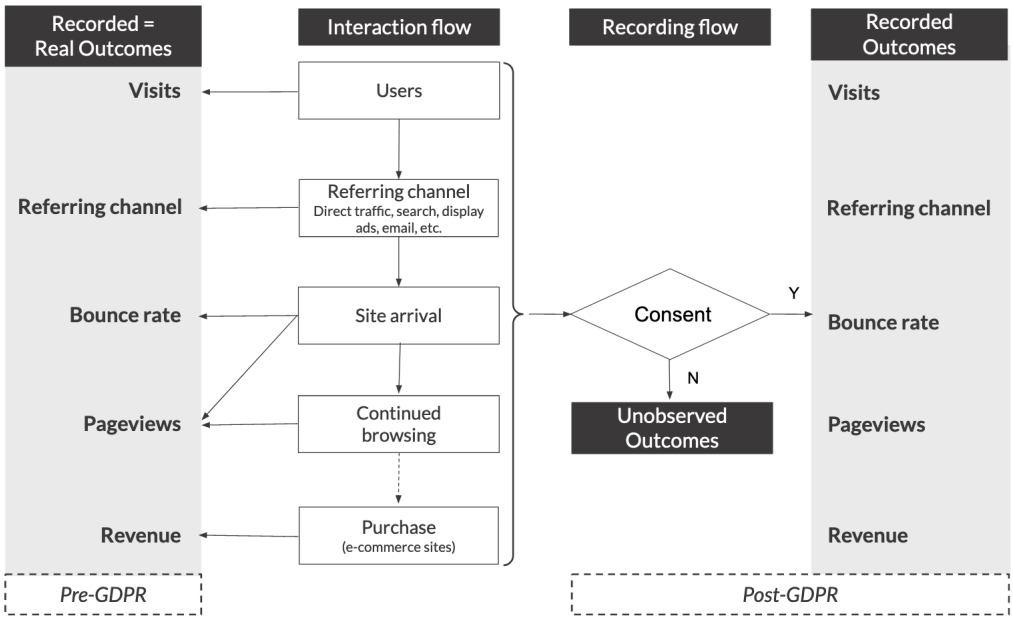
The left hand side of Figure 1 illustrates how web analytics vendors construct economic variables from website traffic pings. When a user arrives on site, a *site visit* is initiated. As we discuss below, sites receive a signal of how the user arrived at the website: the *referring channel*. Once a user arrives on site, the user may either leave the site or continue to browse more site pages. The former case is a referred to as a *bounce* and the share of such visits is referred to as the *bounce rate*. Whether the user bounces or continues to browse onsite, that user generates site *pageviews*. On e-commerce sites, some users complete a purchase and the web analytics vendor will record the site's *revenue*.

One advantage of web analytics data is that they describe how users arrive on site, broken out by referring channel. Sites can coordinate with their analytics vendor to infer these channels from referring links. For instance, the site can send a marketing email to a user containing purpose-built hyperlinks. These hyperlink URLs contain not only the landing page address but also the channel (email) through which the user arrives onsite. We observe four referring channels: search, display advertising, email marketing, and social media. Search includes both paid and natural search, and social media includes clicks on both paid and organic social content. Users can also arrive onsite without clicking a link—e.g. by using a bookmark or by typing the URL into browser navigation bar—which is referred to as the direct navigation channel. Referring channel data can also be used to attribute site outcomes (e.g. pageviews, revenue) to these different marketing channels. These data are therefore also referred to as *last-touch attribution* data, because they record the user's last channel interaction prior to arriving on site. Last-touch attribution is understood to ignore prior channel interactions including those preceding a user's direct navigation to the site. Nonetheless, changes in a marketing channel's last-touch attribution share can indicate a change in channel spending and/or channel effectiveness.

3.2 Data recording pre- and post-GDPR

The GDPR affects the share of web analytics data that vendors, like Adobe, record. Figure 1 emphasizes the difference in data recording before the GDPR (left-hand side) and after the GDPR (right-hand side). Before the GDPR, Adobe records all the user traffic onsite. After the GDPR, Adobe only records outcomes from consenting users. If the user consents, Adobe will observe all that user's outcomes, just as in the pre-GDPR period. If the user revokes consent, Adobe does not record any analytics data associated with that user. As a result, we do not observe any data associated with non-consenting users, including any reference to their visit or site outcomes. We observe *recorded* outcome measures in our data: the

Figure 1: Web analytics data recording before and after the GDPR



consent rate drives a wedge between recorded and real outcomes post-GDPR. To fix ideas, we present a simplified decomposition of our recorded outcomes[8]. Specifically, our recorded outcomes reflect these different components pre- and post-GDPR:

$$\begin{aligned} \text{Pre: } & \textit{Recorded outcomes} = \textit{Real outcomes} \\ \text{Post: } & \textit{Recorded outcomes} = \textit{Real outcomes} \times \textit{Consent rate} \end{aligned} \tag{1}$$

Post-GDPR, recorded outcomes are the product of real outcomes and the consent rate.

The pre- and post-GDPR comparison of recorded outcomes creates an identification challenge. Changes in real outcomes or non-unanimous consent rates (below 100%) both affect recorded outcomes. However, a measured effect on recorded outcomes does not alone reveal each factor's contribution to the change. We expect that the GDPR impacts both a website's real outcomes *and* reduces data recording through consent. Therefore, we must take care in interpreting changes in recorded outcomes. Our analysis thus proceeds in two stages. In Section 5, we measure the effect of the GDPR on recorded outcomes. In Section 6, we propose theory-driven diagnostics and empirically quantify the contribution of each underlying mechanisms, which reflect either consent-related or real outcome-related effects of the GDPR.

[8]Equation 1 is a simplification used to fix ideas, in reality outcomes are likely correlated with consent rates. This would lead to a more complicated functional form for post-GDPR recorded outcomes. We analyze this case in Section 6.

3.3 Sample construction

Our panel data consists of online firms that serve EU residents and that use Adobe Analytics. Our weekly panel data covers 1,084 analytics dashboards for 32 weeks in each of 2017 and 2018.[9] Our unit of analysis is a *dashboard* rather than a website or firm. Firms construct analytics dashboards to provide insight into their online presence, operations, and customers. As such, the dashboard dataset reflects the firm's view of its online presence and can disaggregate large, multi-brand firms. A dashboard may have many-to-many relationships with websites. For example, consider a retailer with both France and UK facing sites. The retailer could choose: 1) to combine both sites into one dashboard, 2) create an dashboard for each site, or 3) create multiple dashboards with duplicate information. In general, firms may elect to organize their websites and/or dashboards by different brands or by customer location. Our primary outcomes of interest are: 1) a pageview: a request for a full webpage document by a site visitor; and 2) revenue: total onsite user spending in US dollars.[10] Both outcomes are economically relevant for online firms. For ad-supported websites, pageviews are proportional to advertising revenue. For e-commerce sites, we directly observe revenue.

Our dashboard sample expands on past work using Adobe Analytics data and undergoes several quality control steps. Drawing from the full sample of Adobe's data, we extract dashboards with at least 500 average weekly visits from the EU prior to the GDPR's enforcement. This threshold ensures EU-relevant data and avoids noisy outcome data due to low EU traffic, while still including long tail sites below the top 100,000 Alexa sites. Some firms may respond to the GDPR's data minimization requirement by dropping Adobe Analytics altogether. To rule this out, we drop dashboards that turn off reporting post-GPDR.[11] We also remove dashboards that are labeled for testing purposes. Furthermore, we remove dashboards whose constituent data elements may have changed as evidenced by unusual growth prior to the GDPR.[12] We filter out dashboards with frequent reporting outages (over 20% of sample weeks), with our resulting panel being 99.98% complete for the pageviews outcome and 99.4% complete for the revenue outcome. Finally, we remove dashboards with outlier user behavior in the pre-GDPR period.[13] Our sample's 353 e-commerce dashboards is sub-

[9]For 2018, our data span the 4th Friday of the year through the week beginning with the 35th Friday of the year (January 28th 2018 to September 7th 2018).

[10]In May 1st, 2018 dollars. Exchange rates are held fixed at May 1st, 2018 rates.

[11]A dashboard may exit our sample because the firm moves its data to a new dashboard, churns from Adobe, or otherwise changes its data structure. In total, less then 0.5% of our dashboards are removed due to this behavior.

[12]For instance, firms may change the list of domains associated with a dashboard. We remove dashboards that grow more than 170% or fall more than 70% between 2017 and 2018 prior to treatment.

[13]Specifically, we remove dashboards with average pageviews per visit and average revenue per visit greater than the 95th percentile of the distribution. At the extreme, these outliers are thirty times the mean pageviews per visit and several thousands of times the mean revenue per visit in Table 1. Note that our main

stantially larger than the sample used in Goolsbee and Klenow (2018) because we include non-US firms, examine more recent data, and do not require product-level data.

Our data offer both economic scale and diversity to investigate the GDPR's impact on the EU's digital economy. The panel includes diverse firms in such industries as finance, health, travel, news/media, and retail. Our sample contains a mix of e-commerce sites, publisher sites, and corporate sites. By observing outcomes by the user's country of origin, we can focus on the EU residents protected by the GDPR. In addition, we gain North American traffic as an alternative control group. Our data contain approximately $0.75 billion per week in European e-commerce revenue, which represents about 12% of total European e-commerce. The data also contain $2.8 billion per week in North American e-commerce revenue—almost a third of North America's estimated spending.[14] The data contain 4.3 billion weekly pageviews from EU users alone. For comparison, Wikipedia saw an estimated 11 billion pageviews from the EU users in a typical month of 2018.[15]

3.4 Summary statistics

Table 1 illustrates the heterogeneity in dashboard sizes in our sample. We report pre-GDPR 2018 summary statistics for pageviews, revenue, usage metrics[16] (pageviews per visit, and revenue per visit), bounce rates, and traffic origin at the dashboard-week level. Dashboards vary in traffic volume from about 7,000 weekly pageviews to almost 5 million weekly pageviews at the 10th and 90th percentiles. The pageview and revenue distributions have long right tails; both means exceed the respective medians by an order of magnitude. This fact motivates our use of logged dependent variables in our analysis. Usage patterns vary by dashboard: on some dashboards users browse less than two pages per visit on average while on others users browse as many as eight. This pattern is evident for the revenue per visit metric as well. Bounce rates are also heterogeneous across sites (10th percentile of 14.5% to 90th percentile of 70.8%) but 40.8% on median. Since traffic from the EU falls under the GDPR's scope, we summarize dashboard EU user outcomes as a share of global outcomes (as well as North America's share for comparison). 41.2% of all dashboards receive more than

effect point estimates in Section 5 do not change substantially with this trimming. However, our pageview and revenue per visit estimates (in Table 5) are imprecise when we include these outliers. Precision aside, these point estimates are robust to different trimming thresholds.

[14]Annual US e-commerce spending in 2017 was $461.5 billion. This is approximately $37.8 billion per month after excluding the large increases in holiday spending in November and December (U.S. Census Bureau, 2019).

[15]https://stats.wikimedia.org/wikimedia/squids/SquidReportPageViewsPerCountryOverview.htm

[16]In principle the usage metrics could be constructed with either visits, or unique visitors, in the denominator. In practice, the distinction between the two is minimal as most visitors only visit a site once within a given week. Empirically, using visits versus unique visitors leads to no real difference in our results.

Table 1: 2018 pre-GDPR weekly summary statistics

Variable	Obs.	Mean	10th percentile	Median	90th percentile
Pageviews (full sample)					
EU user pageviews	1,084	4,008,584	6,722	108,349	4,483,751
EU user pageviews per visit	1,084	4.47	1.73	3.64	7.87
EU user bounce rates	1,084	42.3%	14.5%	40.8%	70.8%
% EU/ Global	1,084	45.1%	0.5%	19.3%	99.1%
% North America/ Global	1,084	39.1%	0.4%	10.8%	97.8%
Revenue (e-commerce sample)					
EU user revenue	353	$505,379	$1,507	$47,872	$1,352,089
EU user revenue per visit	353	$3.35	$0.07	$1.63	$9.26
EU user bounce rates	353	37.1%	16.4%	35.3%	61.3%
% EU / Global	353	46.0%	0.1%	21.3%	99.1%
% North America / Global	353	45.4%	0.2%	22.4%	99.6%

three quarters of their traffic from EU users and 33.7% receive more than three quarters from North America. The sample's traffic tilts towards EU users on average, though the average share of revenue is balanced between EU and North American users. Throughout, the pageviews outcome references the full sample, whereas the revenue outcome references the e-commerce sample.

Table 2 summarizes the mean last-touch attribution channel shares. Table 2 includes the 542 dashboards (and 260 e-commerce dashboards) that collect this data for at least three channels. The last-touch attribution subsample tends to be larger with mean EU weekly pageviews of 8.1 million compared to 4.0 million in the full sample in Table 1. Each dashboard's channel shares are weighted by its pageviews and revenue respectively. The top two channels of direct traffic and search total 91.4% of pageviews and 91.0% of revenue on these sites. Display's referral share is less than 2%, though only 40% of dashboards collect this data for the samples in Table 2. Conditional on reporting, the share of display is 3.3% for all sites and 1.3% for e-commerce. For email, the conditional shares are 5.7% and 7.1% for all sites and e-commerce sites, respectively. By comparison, Budak et al. (2016) report that display ad clicks and email initiate 3% and 7%, respectively, of sessions on the top 10,000 e-commerce sites.

Figure 2 graphs the evolution of log pageviews and of log revenues over time. We use the enforcement of the GDPR—denoted by the vertical line on May 25th, 2018 (week 22)—as an event study in our analysis. Figure 2 also plots same-week 2017 outcomes using a dotted line for comparison. Both pageviews and revenues start off higher in January 2018 than January 2017. The gap is smaller for pageviews, though the two lines cross after the GDPR with 2017 pageviews substantially higher than 2018 by September. Revenue exhibits a larger

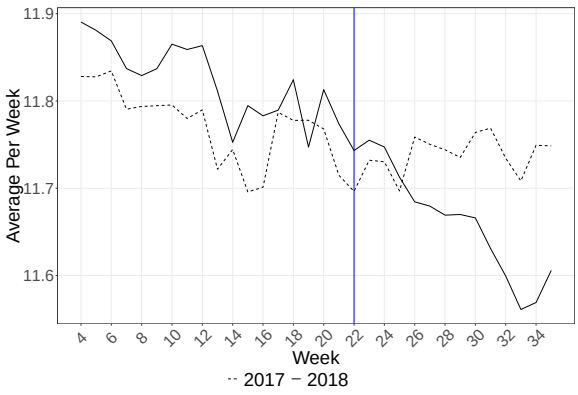
Table 2: Last-touch attribution sample: mean channel shares

	Recorded pageviews	Recorded revenue
Observations	542	260
Direct traffic	33.3%	39.1%
Search	58.1%	51.9%
Display	1.3%	0.4%
Email	3.9%	7.9%
Social	3.4%	0.6%

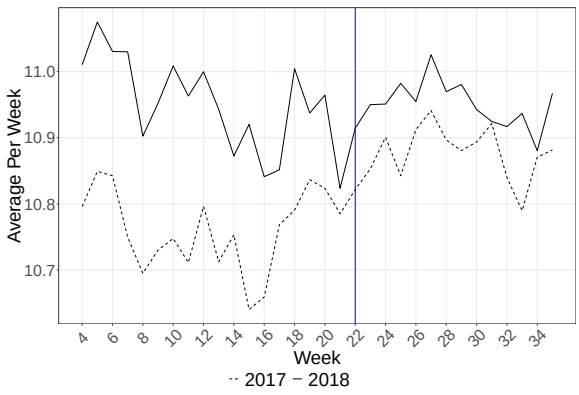
Note: 2018 EU pre-GDPR outcomes for those dashboards that report at least 3 channels. Channels shares within dashboards are weighted by pageviews or revenue respectively.

Figure 2: Evolution of EU user outcomes over time

(a) Log Pageviews



(b) Log Revenue



gap of about 0.2 log points prior to May 25th, but this gap closes substantially afterwards. Both outcomes follow similar trends in 2017 and 2018 prior to May 25th, which motivates our choice of 2017 outcomes as a control group in our empirical strategy below.

4 Empirical strategy

We seek to quantify the impact of the GDPR on recorded website pageviews and e-commerce revenue. In this section, we develop our primary empirical strategy that applies a panel difference estimator approach and uses the prior-year observations for the same firms as a control. We then overview several alternative approaches we pursue for robustness.

4.1 Panel difference estimator (PD)

The GDPR represents a seismic shift in privacy regulation but its scope and imposition creates two challenges for inferring the regulation's effects: 1) identifying an untreated

set of dashboards to serve as a control group; and 2) isolating confounding changes post-enforcement. The GDPR's scope extends beyond the EU's borders so that countries outside of the EU may not represent a clean control group. In particular, its regulations apply to both EU-based firms as well as firms that target EU residents. Table 1 shows that most of our dashboards have a significant share of traffic from the EU. Thus, EU users on non-EU sites are an invalid control group. Furthermore, firms may execute their GDPR compliance strategies across all of their customers in order to reduce customer concerns or the costs associated with treating customers differently by country. Such spillover effects imply that non-EU users are potentially contaminated as a control group. The GDPR's common enforcement date creates a second challenge for inference: post-GDPR changes in outcomes could result from confounding variables that change after May 25, 2018. As an example, in Figure 1 we see that pageviews and revenue fall post-GDPR, but this may be confounded by Europe's summer vacation season. Indeed, Figure 1 reveals that both outcomes fall over over the same period in 2017.

Our control group consists of 2017 outcomes from the same set of dashboards as our treatment group. This ensures that the control group shares seasonal trends and resembles the treatment group. Control units from the past ensures no contamination from the GDPR's enforcement in the EU or its spillover effects beyond the EU. By using the same time period in 2017, this control group accounts for seasonal variation, for instance due to summer vacation and holidays. Further, firms experience different seasonal patterns in demand: surfing websites have more traffic in summer while skiing websites have less. Holding our sample of dashboards constant ensures that the control group captures the seasonal pattern of this specific combination of firms. Figure 1 confirms that pageviews and revenue follow similar trends prior to May 25 in both years. In Appendix B we present placebo tests in which we test for pre-treatment differences between our treatment and control group. For both of our outcomes, no placebo treatment estimate is statistically significant, or of similar magnitude, to our estimated treatment effect in Section 5. The 2017 dashboards are also a natural control group because firm characteristics are relatively fixed year over year. However, we cannot compare the post-May periods between 2018 and 2017 directly because we see in Figure 1 that the pre-May levels are higher in 2018.

Our primary empirical approach applies a panel difference estimator to compare EU user recorded outcomes in 2018 to 2017. The panel difference estimator resembles a difference-in-differences approach[17] in that we assume parallel trends on recorded outcomes as our

[17]Our panel difference estimator is operationalized like a differences-in-differences estimator and relies the same identifying assumptions. The main distinction is that both the control and treatment groups consist of the same set of units (dashboards) though at different times.

identifying assumption. That is, the effect of the GDPR on recorded outcomes is identified by the difference in trends post-May between 2017 and 2018. Recall that all real outcomes are recorded pre-GDPR (including 2017), so any post-GDPR deviation from the counterfactual trend must arise from the GDPR's effect on real outcomes and/or its consent effect data recording (equation 1). Our regression equation is as follows:

$$\log(y_{itw}+1) = \alpha \mathbb{1}\{2018\}_t + \beta\left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post-GDPR}\}_w\right) + \eta_i + \xi_w + \epsilon_{itw} \qquad (2)$$

where i denotes the analytics dashboard, w denotes calendar week, and t denotes year. y_{itw} denotes *recorded* onsite outcomes by EU users. Our model uses log outcomes because the distributions of recorded outcomes are highly skewed (see Table 1). $\mathbb{1}\{2018\}$ is an indicator variable for 2018 observations (treatment) and $\mathbb{1}\{\text{Post-GDPR}\}$ is an indicator variable for calendar weeks after the May 25th enforcement date. We include dashboard- and calendar-week-specific fixed effects denoted by η_i and ξ_ω. Note that we omit the $\mathbb{1}\{\text{Post-GDPR}\}$ term as it is collinear with the calendar-week fixed effects. β represents the estimate of the average effect of the GDPR on EU user recorded site outcomes in our sample.

4.2 Robustness

To address potential threats to validity, we outline alternative empirical strategies. Section 4.2.1, addresses a concern about anticipatory or delayed compliance behavior. Section 4.2.2 outlines a synthetic controls approach that relaxes an underlying restriction in our panel difference estimator and allows us to more flexibly construct a control group. Section 4.2.3, discusses an alternate control group and two associated model specifications that account for any contemporaneous shocks to the web outcomes of Western users.

4.2.1 Window regressions (PD-WR)

To address concerns about anticipatory or delayed compliance by websites, we remove a two-month window around the GDPR from our data and re-estimate the regression in equation (2). Firms made large investments to comply with the GDPR. Surveys reveal that some firms completed this work before the enforcement deadline while other firm's efforts were ongoing (TrustArc, 2018). Since firms can make quick changes to their website and online marketing, they have an incentive to wait until the last minute to implement their online GDPR compliance changes. Other research confirms that most websites waited until the enforcement deadline before making changes to their sites (e.g. Sørensen and Kosta 2019; Johnson et al. 2020). Note also that Figure 2 displays no change in trend before the GDPR

took effect. This is further supported by pre-enforcement placebo checks in Appendix B. Nevertheless, anticipatory or delayed compliance relative to the deadline may lead us to under- or over-estimate the effect of the GDPR in equation (2). We address this by reestimating equation (2) after dropping four weeks of data both before and after the deadline: we term these the "window regression" results.

4.2.2 Synthetic controls (SC)

Our empirical strategy's within-dashboard design uses the 2017 dashboards as the control group to capture firm-specific seasonality and characteristics. However, the difference-in-difference model assumes that the counterfactual trend is best represented by an equal weighting of these dashboards in 2017. Synthetic controls present a data-driven alternative for selecting the control group. We use the synthetic controls approach (Abadie et al., 2010; Doudchenko and Imbens, 2016) to construct a re-weighted control group that relaxes our within-dashboard restriction in order to predict the counterfactual in the post-GDPR period.

Synthetic controls flexibly construct a control group by taking a weighted average of control-units in order to best predict outcomes for treated units. Intuitively, if we can construct a control group that behaves similarly enough to the treatment group in the pre-period, then this control group should behave similarly to how the treatment group would have behaved after the intervention date, had it not received treatment. We define a *control-unit* to be any 2017 dashboard (logged) data and our *treated-unit* to be the mean of our (logged) 2018 dashboard data, plotted in Figure 5. Because we have one control-unit per dashboard (1,084 in our *full* sample and 353 in our *e-commerce* sample) and only one treated unit, we follow Doudchenko and Imbens (2016) and use an elastic net to construct our synthetic control. Weights are chosen in order to minimize the pre-treatment difference between the treated-unit and potential control units. The intent of synthetic controls is to predict the counter-factual, thus we use cross-validation to incorporate prediction error into our objective function. Then, the counterfactual is constructed by taking the chosen weights and using them to aggregate post-treatment control-unit outcomes. We can then recover the treatment effect by differencing the treated and synthetic control outcomes. Appendix A.1.1 details the cross-validation and model fitting procedure.

4.2.3 North American panel difference (PD-NA) and triple panel difference (TPD)

We next consider a contemporaneous control group. Our 2017 control group would not account for any 2018 contemporary confounds like global macroeconomic changes. We thus consider the 2018 web outcomes of North American users from our dashboard sample as an alternate control group. Table 1 reveals that our dashboards have slightly less North American traffic than EU traffic, on average. E-commerce dashboards accrue roughly equal percentages of their revenue from the EU and North America. We explain above that this control group is likely contaminated—owing to within-firm spillovers of GDPR compliance to non-EU users. Thus, any specification with North America as the control group may understate the effect of the GDPR. Other GDPR studies use similar contemporaneous controls (Jia et al., 2018; Aridor et al., 2020; Zhuo et al., 2019), though these authors also acknowledge this contamination issue. We present both our panel estimator with North American controls and a triple panel difference specifications using the North America control. Both specifications address a confounding and contemporaneous shock to online outcomes to both EU and North American users. The triple panel difference specification compares our preferred EU panel difference estimate (equation 2) and an analogous North American panel difference estimate that uses outcomes from 2017 and 2018. The triple panel difference specification identifies the GDPR effect separately from continent-specific seasonality (shared in 2017 and 2018) and a common shock after May 2018 to both EU and North American users.

5 GDPR effect estimates

Table 3 presents the results of our main specification in equation (2). We estimate main effect coefficients of -0.124 for recorded pageviews and -0.142 for recorded revenues. Both estimates are significant at the 1 percent level. To aid interpretation of our non-linear model, we calculate marginal effects (see Appendix C for details). Our point estimates indicate a 11.7 percentage point drop in recorded pageviews and a 13.3 percentage point drop in recorded revenue. For the median dashboard in each sample, this corresponds to a 15,043 drop in weekly recorded pageviews and a $9,227 drop in weekly recorded revenue, respectively. Note that we also provide placebo checks to empirically validate similar pre-trends in Appendix B.

The resulting policy implications vary by whether these estimates reflect changes in real web outcomes, or changes in the share of recorded outcomes. At one extreme, a 12% reduction in real web outcomes would represent a significant economic burden of the GDPR. At the other, a 12% non-consent rate may be construed as progress toward protecting the

Table 3: Panel difference estimator results: 2017 control group

	(1)	(2)
Dependent variable	log(Pageviews + 1)	log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.124	-0.142
	(0.026)	(0.030)
$\mathbb{1}\{2018\}$	0.050	0.193
	(0.014)	(0.028)
Average marginal effect	-11.67%	-13.26%
RSID + Week FE	Y	Y
R^2	0.969	0.962
N	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level, *p<0.1; **p<0.05; ***p<0.01.

data of privacy-sensitive users. The main effect estimates reflect something between these two extremes. Section 6 proceeds to construct bounds on how much the GDPR affects real outcomes versus consent by using theory-driven empirical analysis.

5.1 Robustness

In this section, we present four robustness checks corresponding to the alternate specifications described in Section 4.2: panel differences with North America as the control group (denoted by PD-NA), triple differences (PD-NAEU), window regressions (PD-WR) and synthetic controls (SC). We emphasize the coefficients of interest in Table 4 for clarity.

Table 4 indicates that our main effect results are robust to several alternate specifications. Both specifications using North America as a control group yield smaller point estimates though each remains negative and statistically significant at the 1% level. Of the two specifications, the more conservative marginal effect estimates are -5.4% for recorded pageviews and -5.6% for recorded revenue. We take these lower point estimates as evidence of spillover effects from the GDPR on North American traffic, as discussed in Section 4. In contrast, both our window regression estimates (columns (3) and (7)) are higher than our preferred estimates in Table 3, though they remain within the original confidence intervals. The marginal effects here are -16.0% for recorded pageviews and -16.1% for recorded revenue. These higher estimates could arise if several websites delayed their compliance with the GDPR. Finally, our synthetic control results in columns (4) and (8) indicate marginal effects of -8.7% for recorded pageviews though only -1.4% for recorded revenue. Inference using synthetic controls is difficult: we follow placebo procedures outlined in Abadie et al. (2010). Only 0.2% of placebos achieve a magnitude of prediction error similar to our pageview result.

Table 4: Robustness regression results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Model†	PD-NA	|TPD	PD-WR	SC	PD-NA	TPD	PD-WR	SC
Dependent variable	log(Pageviews + 1)				log(Revenue + 1)			
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\}$		-0.067	-0.174			0.028	-0.176	
		(0.024)	(0.025)			(0.040)	(0.031)	
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post-GDPR}\} \times \mathbb{1}\{\text{EU}\}$		-0.056				-0.149		
		(0.019)				(0.049)		
$\mathbb{1}\{\text{EU}\} \times \mathbb{1}\{\text{Post-GDPR}\}$	-0.084	-0.036			-0.058	0.097		
	(0.005)	(0.002)			(0.013)	(0.010)		
$\mathbb{1}\{2018\}$		0.076	0.056			0.050	0.215	
		(0.017)	(0.015)			(0.037)	(0.028)	
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{EU}\}$		-0.026				0.144		
		(0.016)				(0.041)		
Average marginal effect	-8.09%	-5.44%	-15.98%	-8.70%	-5.59%	-13.87%	-16.15%	-1.40%
RSID + Week FE	N	N	Y		N	N	Y	
RSID + Week + Region FE	Y	Y	N		Y	Y	N	
R^2	0.484	0.491	0.966		0.391	0.397	0.961	
N	64,931	138,280	49,832		20,199	42,721	16,111	

Note: †Model: 1) PD-NA: North American user outcome control; 2) TPD: triple panel differences; 3) PD-WR: window regressions; and 4) SC: synthetic controls. Standard errors clustered at the Dashboard + Week level, *p<0.1; **p<0.05; ***p<0.01.

The revenue synthetic controls estimate is less robust, with 22.4% of placebos achieving a similar magnitude of prediction error. In Appendix A.1 we discuss the synthetic control procedure and results in detail.

Taken as a whole, these alternate specifications seek to address potential threats to validity in our preferred empirical approach, and the results indicate a robust negative impact of the GDPR on recorded site outcomes.

5.2 GDPR effect heterogeneity

We examine whether firm beliefs about regulatory strictness moderate the effect of the GDPR. Firms report heterogeneous investment in and readiness for the GDPR (TrustArc, 2018). Regulators can vary the size of the fines and the probability of levying fines to induce firms to comply with regulation. As such, firms that face a more strict regulator may work harder to comply with the law. For instance, these firms may cut back on personalized marketing or set a higher bar for collecting consent on their website. This exercise is interesting from a regulatory design perspective and provides additional evidence that the GDPR explains the changes in EU traffic after May 25, 2018.

To examine the role of regulatory strictness, we exploit geographic variation in firm's be-

liefs about their country's data protection authority. Though the GDPR harmonizes privacy regulation in the EU, the GDPR's enforcement is split between a central EU supervisory authority and authorities in each EU country. Firms are assigned to a country's supervisory authority based on where their customers or headquarters are located. To quantify regulatory strictness by EU country, we use of a measure from the European Commission (2008) survey of 4,835 data controllers. Controllers report whether data protection law is "interpreted and applied more rigorously" in their country than in the rest of the EU. We construct a normalized index that ranges from -1.64 in Greece to 1.49 in Sweden.[18] We assign each dashboard to an EU country based on the member state that drives the majority of its web traffic in 2018 before the GDPR.

Our resulting estimates reject the null hypothesis that regulatory strictness is unrelated to the GDPR's effect on recorded web outcomes. We reestimate our panel differences model (equation 2) further including an interaction with regulatory strictness. We also include another interaction term for income per capita, which may confound the role of regulatory strictness. We present the estimates in Table 7 of Appendix D.1. We estimate a regulatory strictness interaction coefficient of -0.040 log points (s.e. 0.006) for pageviews and -0.041 (s.e. 0.019) for revenue, which are both significant at the 1% level. The estimates imply that a one standard deviation increase in regulatory strictness, holding all else fixed, reduces recorded pageviews by -2.1% and recorded revenue by -4.5% for the respective median dashboards. Thus, country-level differences persist in practice, despite the GDPR's promise of a consistent regulatory environment within the EU. The moderating role of regulatory strictness is a robust finding of the GDPR literature. Johnson et al. (2020) also show that regulatory strictness is correlated with website cuts to their tech vendor use post-GDPR. Jia et al. (2018) show that regulatory strictness is correlated with post-GDPR reductions in EU tech venture investment. These findings further substantiate that our main effect estimates capture the impact of the GDPR rather than an unrelated post-enforcement shock to EU web traffic.

6 Disentangling real and recorded outcomes

To better evaluate the impact of the GDPR on the online economy, we need to separate the measured effect on recorded outcomes from its effect on real outcomes (equation 1). We begin by presenting a decomposition of our recorded outcomes that highlights key features that may be impacted by the GDPR. We then lay out three principle mechanisms that

[18] After excluding non-responses, we construct a country-level index taking values from 0 (all responses are 'totally disagree') to 1 (all responses are 'totally agree'). We normalize the index so that it has a mean of zero and standard deviation of 1.

can explain the GDPR effect: consent, marketing, and privacy frictions. In Sections 6.1 to 6.3, we derive empirical tests for each mechanism that allow us to quantify their individual contribution to the overall GDPR effect.

We begin by decomposing each of our recorded outcomes into its intensive and extensive margins:

$$E[Y] = E[y] \cdot E[N]$$

where Y is our recorded outcome variable (pageviews or revenue) in levels, y is a *usage metric* (average pageviews or revenue per visit), and N is the number of recorded visits. This decomposition relies on an independence assumption between site visits (N) and usage (y) across websites.[19] We emphasize that the GDPR's first order effect is on recorded visits whereas its effect on usage is second order.[20] We assume that the pre- ($t = 0$) and post- ($t = 1$) GDPR recorded visits are related as follows:

$$E[N|t = 1, c = 1] = E[N|t = 0] \cdot (1 - \delta) \cdot (1 - \theta)$$

where $E[N|t = 0]$ is the average baseline number of (real) visits. c is an indicator for consent, which clarifies that post-GDPR recorded visits are conditional on consent. θ captures the consent effect of the GDPR on data recording (missingness) as $\theta \geq 0$ is the non-consent rate: the share of visits where the user does not consent to data collection post-GDPR. δ captures the real effect of the GDPR (on the share of users who provide consent), where $\delta \geq 0$ is the rate at which real visits fall post-GDPR. Post-GDPR, we do not observe non-consenting users, instead we focus attention on measuring effects to the recorded outcomes of consenting users. To do this, we restrict the consent and real effects on recorded visits using a multiplicative functional form assumption.

Under the independence and functional form assumptions, recorded outcomes in the pre- ($t = 0$) and post- ($t = 1$) GDPR periods can be written as:

$$\begin{aligned} E[Y|t = 0] &= E[y|t = 0] \cdot E[N|t = 0] \\ E[Y|t = 1, c = 1] &= E[y|t = 1, c = 1] \cdot E[N|t = 0] \cdot (1 - \delta) \cdot (1 - \theta) \end{aligned} \tag{3}$$

In our data, we observe each of the conditional expectations in equation (3) above. However, we cannot separately identify the number of *real* visits $E[N|t = 0] \cdot (1 - \delta)$ and the consent

[19] Under independence, we have $E[Y] = E[y \cdot N] = E[y] \cdot E[N]$. The independence assumption is strong, but since GDPR precludes the ability to directly observe consent rates, strong assumptions are required to separately identify consent-related censoring of outcomes from changes to outcomes within consenting and non-consenting user groups.

[20] This is validated empirically. Our visits estimates are close in magnitude to our main effect estimates. We estimate a marginal effect of -11.99 for the visits outcomes.

rate $(1-\theta)$ without additional assumptions. This is the key identification challenge of this paper. We can relate the above decomposition to the percentage marginal effect of the GDPR that we estimated in Table 3 as follows:

$$\frac{E\left[Y|t=1,c=1\right]-E\left[Y|t=0\right]}{E\left[Y|t=0\right]}=\frac{E\left[y|t=1,c=1\right]\cdot E\left[N|t=0\right]\cdot(1-\theta)\cdot(1-\delta)}{E\left[y|t=0\right]\cdot E\left[N|t=0\right]}-1$$

Solving for θ, we have

$$\theta=1-\frac{E\left[y|t=0\right]}{(1-\delta)\cdot E\left[y|t=1,c=1\right]}\left(1+\frac{E\left[Y|t=1,c=1\right]-E\left[Y|t=0\right]}{E\left[Y|t=0\right]}\right) \tag{4}$$

In words, we can relate the consent (θ) and real effects (δ) of the GDPR to the marginal effect $\left(\frac{E[Y|t=1,c=1]-E[Y|t=0]}{E[Y|t=0]}\right)$ of the GDPR on recorded outcomes given the pre-and post-GDPR recorded usage ratio $\frac{E[y|t=0]}{E[y|t=1,c=1]}$. This relationship holds due to an adding-up constraint.

In evaluating potential mechanisms for the GDPR estimates in Section 5, equation (3) clarifies the primary channels through which our candidate mechanisms may impact recorded outcomes. Our usage metrics reflect the behavior of recorded individuals pre- and post-GDPR. To the extent that there is user selection on consent or privacy frictions, we will find $E\left[y|t=1,c=1\right]\neq E\left[y|t=0\right]$. We expect GDPR consent dialogs may introduce frictions that reduce user browsing, which puts downward pressure on $E\left[y|t=1,c=1\right]$. On the other hand, consenting users are likely to browse more per visit than non-consenting users, leading recorded usage $E\left[y|t=1,c=1\right]$ to increase relative to the pre-GDPR baseline. Finally, we examine if the GDPR may reduce the effectiveness of personalized marketing, which we expect would reduce site visits and imply $\delta>0$. Below, we discuss these three mechanism in turn, then summarize how each contributes to the intensive and extensive margins in equation (3).

Consent User consent is the primary basis for processing web analytics data. As Figure 1 and equation (3) suggest, if the site collects and respects user consent, the share of recorded data will be a function of the share of consenting users. During our sample period, the majority of websites that served EU users relied on *de facto* opt-out consent. Investigations during that time reveal that most websites did not wait for consent before interacting with third party domains (Johnson et al., 2020) or setting identifier cookies (Sanchez-Rola et al., 2019). In 2019, the Irish privacy regulator found that almost all sites it investigated were still setting cookies upon the user's arrival (Data Protection Commission, 2020b) and the Dutch privacy regulator found most sites it investigated improperly relied on opt-out consent (Autoriteit Persoonsgegevens, 2019). This approach to consent ensures high consent rates,

and the largest GDPR consent management platform reported average consent rates in excess of 90% (Quantcast, 2018). By contrast, sites that followed a strict opt-in approach have much lower consent rates. In this case, the British privacy regulator's site and the Dutch public broadcaster's sites report consent rates of 10% or lower (Information Commissioner's Office, 2019; Snelders et al., 2020). The former released Google Analytics data before and after this change, which revealed that recorded outcomes fell over 90%. We focus on measuring the average non-consent rate across our dashboards to contribute to the policy discussion on *de facto* consent under the GDPR. In Section 6.2, we build on equation (4) using a simple self-selection argument: we expect users who consent to data collection to be users who value the site more, suggesting usage metrics should rise post-GDPR. This approach allows us to obtain empirical bounds on average non-consent rates.

Marketing The GDPR may hurt a firm's real outcomes by reducing the quantity and effectiveness of its marketing. Firms report high costs of complying with the GDPR, which may divert funds from discretionary expenses like marketing. The GDPR also raised the legal risk and logistical cost associated with personal data processing. Firms may respond by reducing their investment in marketing channels like e-mail and online display advertising. Those media rely on personal data in the form of cookie-based identifiers and e-mail addresses respectively. For instance, many firms sent permissioning emails prior to the GDPR seeking opt-in consent to continue emailing users, then dropped non-consenting users from email marketing lists. de Matos and Adjerid (2019) indicate that only 64% of consumers provided such consent even when offered incentives, but that the firm could effectively market to those consumers who consented. The quantity and effectiveness of personalized marketing channels may thus fall post-GDPR. This may be an important effect, as Table 2 suggests that e-mails precede 3.9% of pageviews and 7.9% of revenue to e-commerce sites while display ads precede 1.3% of pageviews and 0.4% of e-commerce revenue. In Section 6.3, we combine heterogeneity in treatment effects across last-touch attribution channels with additional assumptions on equation (4) to identify the marketing component of the real GDPR effect δ.

Privacy frictions Websites are reluctant to use obtrusive consent dialogs that add friction to the user's browsing experience. The Irish data regulator notes a "general resistance" among sites to introduce privacy frictions through the consent interfaces (Data Protection Commission, 2020b). For this reason, websites experiment with different consent interfaces to reduce privacy frictions and ensure a high consent rate (Long, 2020). Past research suggests that interruptions can hurt online usage (Lambrecht et al., 2011) and noted the long time required to read websites privacy policies (McDonald et al., 2009). Thus, the

specific concern of privacy frictions is that they interrupt the user's browsing and deter users from continuing to browse the site - leading to a decrease in real outcomes. This motivates a simple empirical test: we expect the share of visits where the user *bounces* (browses a single page before leaving) will increase as privacy frictions increase. In Section 6.1, we test for changes in the bounce rate after the GDPR.

The three mechanisms above are key for the GDPR, though other real effect mechanisms may also be present. For example, users may change the sites they browse and their browsing intensity as the GDPR makes both data collection and data protection more salient. As such, we bound the roles of recording (consent) and the real effects (privacy frictions and marketing) of the GDPR, before pinning the latter down using stronger assumptions related to these mechanisms.

Returning to equation (3), we summarize the extensive and intensive margin effects of the GDPR in the table below. The consent effect reduces recorded visits while marketing and privacy frictions reduce real visits to the site—captured by θ and δ respectively. We expect that consent will increase average usage y due to the favorable selection of consenting users. Finally, we assume that the real effect of the GDPR *only* affects the site usage metrics through the privacy frictions effect on bounce rates.

	Consent(Recording effect)	Marketing & Privacy Frictions (Real effect)
Recorded visits (N)	$\downarrow$ (θ)	$\downarrow$ (δ)
Usage per visit (y)	$\uparrow$	$\downarrow$ *privacy frictions only

We proceed to detail empirical tests for each mechanism and quantify the contribution of each to the overall estimates from Section 5. To build the empirical argument, we consider the mechanisms in the following order: privacy frictions, consent, and marketing. In Section 6.1, we do not find evidence in support of the privacy frictions mechanism and proceed by assuming that inter-temporal changes in usage metrics arise solely from consent-based selection. Section 6.2 shows favorable selection in the usage of consenting users, and bounds the consent and real (broadly-defined) effects of the GDPR. Section 6.3 examines changes in visits from personalized marketing channels to quantify the real effect of the GDPR via marketing.

6.1 Privacy frictions: bounce rate evidence

We test for a privacy frictions mechanism by examining the effect of GDPR on bounce rates. Bounce rates are defined as the share of site visits with only a single pageview. More generally, bounce rates are a key diagnostic outcome in site analytics. Sites may wish to

reduce bounce rates in order to increase ad and e-commerce revenue opportunities. On the other hand, high bounce rates can also indicate that the website effectively communicates information to the user: e.g. the temperature in Paris. In the pre-GDPR period, bounce rates average 42% across all dashboards and 37% in the e-commerce sample.

Bounce rates are useful because they indicate how dashboards seek consent. We expect that bouncing users are unlikely to provide explicit consent, because these users minimally interact with the site. If sites employ opt-out consent, then we expect bounce rates to rise somewhat if the privacy frictions mechanisms holds. If instead the site uses a strict opt-in model, then we expect to see a large drop in bounce rates because we expect few bouncing users will opt in to data recording. Comparing individual dashboard means across the pre- and post- GDPR periods suggests that the large majority of our sites see minimal changes to their bounce rates. In particular, only 3 sites exhibit patterns that are consistent with a strict, opt-in approach in that their bounce rates decrease more than 20 percentage points, but their recorded pageviews fall more than 50%. Thus, the vast majority of websites in our sample appear to employ an opt-out approach for consent, which is consistent with large surveys of website behaviors during that period (Johnson et al., 2020; Sanchez-Rola et al., 2019). As such, consistent with opt-out consent, we expect bounce rates to rise if privacy frictions are contributing substantially to our point estimates. We use bounce rates to test for the privacy frictions mechanism whereby GDPR consent dialogs dissuade users from further browsing.

We test for a change in bounce rates by reestimating our panel differences model (equation 2) with bounce rates as the dependent variable. Table 5 columns (1) and (2) present the model estimates: -0.275 percentage points (s.e. 0.304) for all sites and -0.354 percentage points (s.e. 0.591) for e-commerce sites. Thus, we find no statistically significant evidence that bounce rates change due to the GDPR and these null effect estimates are precisely estimated. This finding may allay website concerns about the privacy frictions mechanism after the GDPR, though more obtrusive consent dialogs that seek opt-in consent could create greater browsing frictions. We proceed under the assumption that the privacy frictions mechanism does not materially contribute to our estimated GDPR effect.

6.2 Consent: site-usage metric evidence

We next leverage a self-selection assumption to bound the contribution of consent to our GDPR estimates. We assume that users' site browsing intensity and willingness to consent are *correlated*. Changes in browsing intensity metrics recorded by Adobe Analytics then indicate selection by users who consent as well as the loss of users who do not consent. We

Table 5: User-driven mechanism results: bounce rate & site usage outcomes

	Bounce rate		*Usage metrics*	
	(1)	(2)	(3)	(4)
Sample	All Dashboard	E-commerce	All Dashboard	E-commerce
Dependent variable	Bounces per visit	Bounces per visit	Pageviews per visit	Revenue per visit
$\mathbb{1}\{EU\}\times \mathbb{1}\{Post\ GDPR\}$	-0.275	-0.354	0.200	0.172
	(0.304)	(0.591)	(0.071)	(0.093)
$\mathbb{1}\{2018\}$	0.311	0.111	-0.291	0.041
	(0.304)	(0.539)	(0.067)	(0.094)
Implied θ bounds	N/A	N/A	[4.08%, 15.36%]	[5.26%, 17.69%]
RSID + Week FE	Y	Y	Y	Y
R^2	0.861	0.790	0.724	0.758
N	69,344	22,436	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level, *p<0.1; **p<0.05; ***p<0.01

expect positive selection due to consent: that is, users who value the site more will both browse the site more and will be more willing to provide consent. We test this using two browsing intensity metrics which we refer to as *usage metrics*: pageviews per visitor and revenue per visitor. We again apply our panel differences specification (equation 2), with each usage metric as the dependent variable. Conceptually, we obtain a lower bound by noting that observed usage falls by the product of the non-consent rate and the usage rate among consenting users. We note that the usage of non-consenting users has a natural lower bound of one pageview per visitor. This implies a lower-bound for the non-consent rate. We then obtain an upper bound by assuming the full observed effect on traffic arises from the consent effect on data recording.

Table 5 presents our site usage metric results in levels. We see a statistically significant increase in pageviews per visitor of 0.200 (s.e. 0.071) and a marginally significant increase in revenue per visitor of \$0.172 (\$0.093).[21] Both estimates are positively signed, which suggests positive selection of consenting users post-GDPR. Our average dashboard sees 4.47 pageviews per visit in the pre-GDPR period, so our estimate represents a relative increase of 4.5% in pageviews per visit post-GDPR. For e-commerce dashboards, our point estimate is noisier but corresponds to an increase in revenue per visit of 5.1%. These findings accord with Aridor et al. (2020), who also show favorable selection among consenting users post-GDPR.

[21]Note that we also estimate the GDPR effect on pageviews per visitor in our e-commerce sample, which yields a positive and statistically significant estimate of of 0.213 (0.079).

6.2.1 Consent effect identification

We formalize the selection argument above in order to bound the consent effect. We leverage the fact that the intensive margin—given by the usage metrics—is invariant to the GDPR's real effect on visits. Having ruled out privacy frictions, we can write the site usage metrics as a mixture of usage from consenting and non-consenting users. Prior to the GDPR, average site usage is a weighted average of: 1) the usage of user visits that *would consent* under the GDPR and 2) the usage of user visits that *would not consent.* After the GDPR, site usage only reflects the visits of those users who consent. Let y denote the average usage outcome, c be an indicator for consent, and θ denote the non-consent share as before. Then, the expected usage in the pre- ($t = 0$) and post- ($t = 1$) GDPR periods can be written as follows:

$$E[y|t=0] = (1-\theta)E[y|c=1,t=0] + \theta E[y|c=0,t=0] \tag{5}$$

$$E[y|t=1] = E[y|c=1,t=1] \tag{6}$$

Our consent effect identification approach relies on the assumption that the usage among users who would consent is constant before and after the GDPR: i.e. $E[y|c=1,t=0] = E[y|c=1,t=1]$. Under this assumption, usage among consenting users is empirically identified by $E[y|t=1]$. Note that this assumption rules out a real GDPR effect on usage among consenting users. The validity of that assumption is strengthened by our above finding that the GDPR did not create privacy frictions that affected the bounce rate.

Solving for θ, we have:

$$\theta = \frac{E[y|t=1,c=1] - E[y|t=0]}{E[y|t=1,c=1] - E[y|c=0,t=0]} \tag{7}$$

In other words, the opt-out rate is the ratio of the GDPR effect on recorded usage and the difference in usage among users who do or do not consent ($E[y|c=1,t=0] - E[y|c=0,t=0]$). We assume consenting users are selected in that $E[y|c=1,t=0] \neq E[y|c=0,t=0]$, an assumption that is supported by our results in Table 5. Equation (7) relates θ to the observable pre-and post-GDPR usage as well as the unobservable $E[y|c=0,t=0]$. Below, we use equations (4) and (7) to construct bounds for θ. Given these bounds, we can use equation (4) to map out the identification region for θ and δ.

Lower bound Equation (7) implies an increasing relationship between θ and the unobservable $E[y|c=0,t=0]$. However, this unobservable has the lower bound $E[y_{pv}|c=0,t=0] \geq 1$ for the pageviews-per-visit outcome (y_{pv}), because each visit mechanically has at least one

pageview. Thus, the lower bound for θ_{pv} is given by:

$$\underline{\theta}_{pv} = \frac{E\left[y_{pv}|t=1,c=1\right] - E\left[y_{pv}|t=0\right]}{E\left[y_{pv}|t=1,c=1\right] - 1} \tag{8}$$

In other words, the consent rate must be as least as large as implied by the maximal difference in usage between consenting users and (unobserved) non-consenting users. For the revenue-per-visit outcome (y_{rv}), the unobservable has the lower bound $E\left[y_{rv}|C=0,t=0\right] \geq 0$, because visits should at worst generate no revenue. Thus, the lower bound for θ_{rv} is given by:

$$\underline{\theta}_{rv} = \frac{E\left[y_{rv}|t=1,c=1\right] - E\left[y_{rv}|t=0\right]}{E\left[y_{rv}|t=1,c=1\right]} \tag{9}$$

Upper bound Equation (7) does not imply an informative upper bound for θ because the equation implies that $\theta \to 1$ as $E\left[y|c=0\right] \to E\left[y|t=0\right]$. Instead, we return to the adding-up constraint imposed by equation (4). If we assume the GDPR did not benefit real site traffic ($\delta \geq 0$), then the consent effect is largest when the real effects is smallest ($\delta = 0$). The upper bound for θ given by equation (4) is then:

$$\overline{\theta} = 1 - \frac{E\left[y|t=0\right]}{E\left[y|t=1,c=1\right]}\left(1 + \frac{E\left[Y|t=1,c=1\right] - E\left[Y|t=0\right]}{E\left[Y|t=0\right]}\right) \tag{10}$$

In other words, this upper bound assumes that the entire GDPR effect comes from consent.

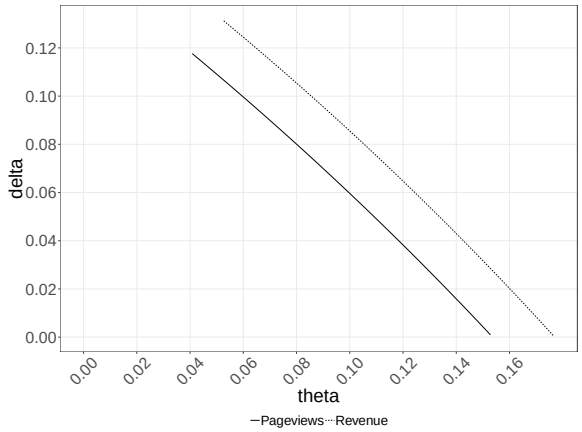
6.2.2 Consent effect bounds

We use our point estimates to construct bound estimates for the share of non-consenting users.[22] Table 5 includes a row with the resulting bounds from equations (8)-(10). Note that these bounds are in terms of visits. In column (3), the non-consent rate (θ) bounds for the full sample are between 4.1% and 15.4%. The revenue per visit estimate delivers tighter bounds on the non-consent rate for e-commerce dashboards: between 5.3% and 17.7%. These bounds are consistent with the over 90% GDPR consent figures recorded by the largest GDPR consent dialog provider (Quantcast, 2018).

Under these assumptions, the bounds on consent have consequences for the size of the real effect (δ) of the GDPR on visits. Equation (4) specifies the functional relationship between θ and δ. Figure 3 graphs this relationship. The consent effect's lower bounds imply an upper bound on the real effect δ: the real effect can be at most 11.7% for the full sample

[22]To operationalize this, we predict $E\left[y|t=1\right] - E\left[y|t=0\right]$ (holding the year fixed at 2018) using our regressions estimates from Table 5.

Figure 3: Identification region for the consent effect (θ) and real effect (δ) of the GDPR



and at most 13.1% for the e-commerce sample. Recall that these bounds and Figure 3 rely on point estimates and ignore the precision of these estimates for the sake of brevity.

For greater economic interpretability, we convert these visit-level consent bounds into pageview and revenue terms. For all sites, the pageview-level non-consent rate bounds are [0.8%, 11.7%]. The pageviews upper bound is given by the full GDPR estimate on recorded pageviews from Table 3. The lower bound of 0.8% is conservative because we assume pageviews per visit for non-consenting users is equal to one. Thus, the lower bound is lower than the visits level bound (4.1%) because non-consenting users have lower usage. The equivalent revenue-level non-consent bounds [0%, 13.3%] are not informative. The upper bound is the recorded revenue estimate from Table 3. However, the lower bound is uninformative because it conservatively assumes the lower bound $E[y|C = 0] = \$0$; in other words, the non-consent users have no revenue impact because they are assumed to generate no revenue at this lower bound. Thus, consent accounts for at least 7% of our overall GDPR effect estimates on pageviews for all sites, which we calculate by dividing the level effect at the lower consent bound by the corresponding main effect point estimate. In place of an equivalent revenue-level figure, we note that the lower bound on visits for e-commerce sites is 29% of the upper bound that assumes all of the GDPR effect is due to consent. We conclude that the effect of the GDPR on data recording—via user consent—contributes importantly to our GDPR estimates on recorded site outcomes.

Consent rates can have real consequences for these firms. First, online firms can use analytics data to improve business outcomes (Berman and Israeli, 2020), but firms will have both less data and selected data post-GDPR. Second, many websites rely on revenue from personalized advertising that relies on third-party cookies. Two recent studies on the value of personalized advertising both found that ad prices fall by 52% without personalization (Johnson et al., 2020; Ravichandran and Korula, 2019), while the Competition and Markets

Authority (2020) report suggests this figure is conservative. We expect non-consent rates for personalized advertising to be at least as high as non-consent for site analytics. Vendors were more likely to rely on consent (rather than "legitimate interest") for advertising than for site measurement purposes (Matte et al., 2020). EU regulators have long viewed personalized advertising with greater concern and its consent requirement as a more clear-cut case (Article 29 Data Protection Working Party, 2012). Combining the 52% figure with our pageview-level non-consent bound estimates, we obtain that websites with advertising would lose between 0.4% and 6.1% in ad revenue due to non-consenting users.

6.2.3 Consent & site size

Sites may vary in their ability to gain consent. For instance, theoretical work by Campbell et al. (2015) suggests that large sites may have superior ability to gain consent, so that privacy consent can create anti-competitive consequences. We first look for differences in the overall GDPR effect by site size by adding an interaction term for firm size to our main panel difference model (equation 2). Our site size indicator splits the (respective) samples by whether they are above or below the median firm. Our results in Table 8 of Appendix D.2 suggest that large e-commerce sites exhibit a lower GDPR effect than small e-commerce sites, though we see no significant differences in our full sample regression for recorded pageviews. Our estimates imply that recorded revenue falls by 8.9% for large sites and 17.4% for small sites, and that difference is marginally significant. Moreover, Table 9 of Appendix D.2 reveals important differences in revenue-per-visit between large and small sites. These revenue differences imply different non-consent bounds. Small firms have non-consent rates between 6.9% and 25.5%, while large firms have lower non-consent bounds between 2.3% and 9.2%. This result provides some empirical support—at least for e-commerce dashboards—that smaller firms are less able to get consent.

6.3 Marketing: last-touch attribution evidence

The 542 dashboards in our sample that track last-touch attribution data tend to be larger, but exhibit approximately the same GDPR effect. Online firms choose whether to track last-touch attribution data using Adobe Analytics, so the last-touch attribution sample is selected (see Section 3.4). The last-touch attribution dashboard sample has more average weekly pageviews (8.1 million) than the average dashboard in our full sample (4.0 million). These firms are also be selected in that they may be more sophisticated marketers or may be more reliant on marketing to generate site traffic. We first reestimate our main specification (equation 2) on the last-touch attribution sample. On the last touch attribution sample, we

estimate a GDPR effect of -0.07 (s.e. 0.022) log points for pageviews with a corresponding marginal effect of -7.1% and -0.09 (s.e. 0.038) log points for revenue with a marginal effect of -8.7%. Our marginal effect estimates correspond to a decrease of -10,635 pageviews per week for the median firm in our last-touch attribution sample and -$7,686 per week in revenue for e-commerce firms. For pageviews, our subsample estimates are statistically different from the full sample estimates, though the 95% confidence intervals overlap. For revenue, our subsample estimates are not statistically significantly different from full sample point estimates, where our marginal effects were -13.3% (Table 3).

We next estimate marketing channel-specific GDPR effects in Section 6.3.1 and use this to estimate the contribution of the marketing mechanism to our overall GDPR effects in Section 6.3.2.

6.3.1 Regression results

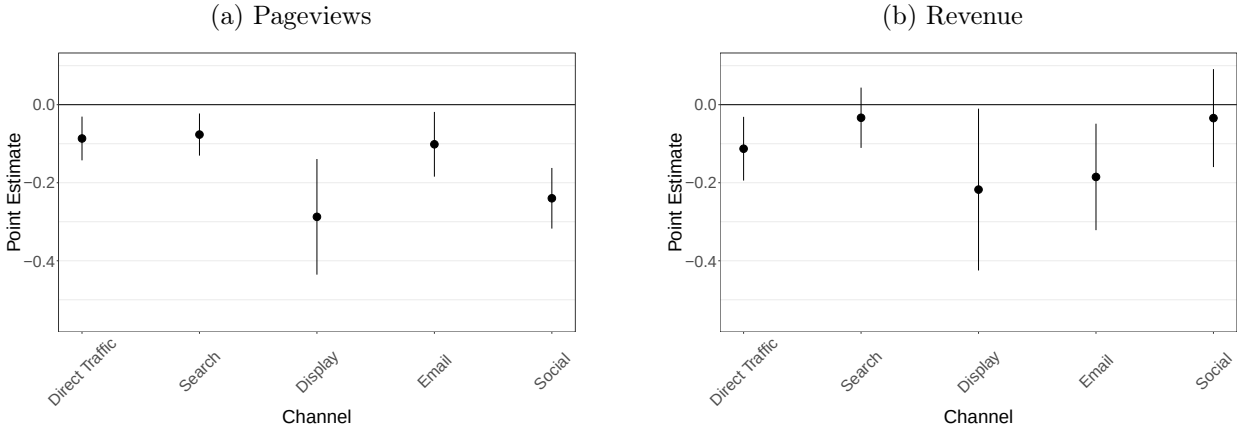
Section 3 highlights how user traffic can arrive on site from different sources. Heterogeneity in the impact of the GDPR across these sources may indicate an effect on the GDPR on marketing. The following regression equation estimates channel-specific GDPR effects:

$$\begin{aligned}\log\left(y_{itwc}+1\right) = &\sum \alpha_c \left(\mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \\ &\sum \gamma_c \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \\ &\sum \beta_c \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{\text{Channel}\}_c\right) + \eta_{ic} + \xi_w + \epsilon_{itwc}\end{aligned} \tag{11}$$

where c refers to attribution channel, w to calendar week, t to year and i to dashboard. y_{itwc} denotes recorded outcomes associated with channel c. η_{ic} and ξ_ω are dashboard-channel and week-specific fixed effects, respectively. We interact an indicator for channel c, $\mathbb{1}\{\text{Channel}\}_c$, with each term of equation (2). β_c are the coefficients of interest because they capture the effect of the GDPR by channel c. Under the specified model, the variation identifying each β_c will be differences in outcomes across channels and years in post-May 25 weeks, after accounting for unobservables common to dashboard-channel and week observations, and a common level shift in channel outcomes across years.

Figure 4 presents the results of estimating equation (11). Channels are labeled on the x-axis. Figure 4 includes 95% confidence intervals constructed from standard errors clustered at the dashboard-week level. As discussed in Section 3.4, firms choose whether to collect last-touch attribution data for individual channels. Each channel treatment effect estimate is therefore conditional on the set of dashboards that collect data for that channel. Channel-specific sample size variation is responsible for the large differences in the width of the

Figure 4: Last-touch attribution treatment effect estimates by channel



confidence intervals presented in Figure 4.

Panel (a) of Figure 4 presents the channel-level marginal effect estimates for pageviews and shows a -8.3% drop in direct traffic due to the GDPR.[23] The more negative and statistically significant impact on paid channels stands out: display (-25.0%) and email (-9.7%). This finding is consistent with the GDPR's protection of personal data and EU regulators focus on both display and email advertising. We see a negative impact on search (-7.4%) and social advertising (-21.4%) as well, which is somewhat counter to our expectations that these channels could compensate for reductions elsewhere. This result could arise from changes in advertising practice, such as cutting ad budgets or reduced targeting effectiveness on these platforms. These estimates suggest that the GDPR adversely affected website marketing activities. In contrast to paid channels, our point estimates for direct traffic and search are smaller in magnitude.

Panel (b) of Figure 4 presents the channel-level treatment effect estimates for e-commerce revenue. Revenue from the direct traffic channel falls -10.7% due to the GDPR, which resembles the effect on pageviews from direct traffic. The paid channel revenue effects also resemble their pageview counterparts, with revenues originating from display advertising falling -19.6% and revenues originating from the email channel declining by -16.9%. Unlike pageviews, we find no significant GDPR impact on revenues originating from other channels including paid search and social ads. These findings are consistent with e-commerce sites moving their ad budgets away from channels that use personal data (display and email) towards those that do not (search and social).

[23] Regression estimates of β_c are included in Appendix E

6.3.2 Marketing effects

While the point estimates in Section 6.3.1 are suggestive of a marketing effect they do not disentangle consent from marketing. To accomplish this, we extend our model from Section 6.2 to accommodate multiple channels. With additional assumptions, we can isolate the the GDPR's marketing effect on consenting users[24]. For ease of exposition, we illustrate this using a two channel example with a direct channel (denoted without primes) and a personalized marketing channel (denoted using primes). To begin, note we can write equations (3) by channel:

$$\begin{aligned} E\left[Y|t=0\right] &= E\left[y|t=0\right] \cdot N_0 \\ E\left[Y|t=1,c=1\right] &= E\left[y|t=1,c=1\right] \cdot N_0 \cdot (1-\delta) \cdot (1-\theta) \\ E\left[Y'|t=0\right] &= E\left[y'|t=0\right] \cdot N_0' \\ E\left[Y'|t=1,c=1\right] &= E\left[y'|t=1,c=1\right] \cdot N_0' \cdot (1-\delta') \cdot (1-\theta') \end{aligned} \tag{12}$$

We make two key two assumptions: (1) consent rates are the same across channels ($\theta = \theta'$), and (2) marketing has no effect on direct traffic ($\delta = 0$). In combination, these assumptions pin down the non-consent rate from the treatment effect estimate for the *direct traffic* attribution channel. Both assumptions are strong. Assumption (1) rules out user selection in consent rates by channel; for example, lower consent rates in channels that attract lower quality users. Assumption (2) rules out spillovers from the firm's marketing to its direct channel. We may expect marketing to increase direct traffic so that our resulting bounds may underestimate the role of marketing.

Under assumptions (1) and (2), θ is identified by the direct traffic channel:

$$\theta = 1 - \frac{E\left[y|t=0\right]}{E\left[y|t=1,c=1\right]} \left(1 + \frac{E\left[Y|t=1,c=1\right] - E\left[Y|t=0\right]}{E\left[Y|t=0\right]}\right) \tag{13}$$

Given θ and and substituting $\theta = \theta'$, we solve for the personalized marketing effect δ' as follows:

$$\delta' = 1 - \frac{1}{(1-\theta)} \frac{E\left[y'|t=0\right]}{E\left[y'|t=1,c=1\right]} \left(1 + \frac{E\left[Y'|t=1,c=1\right] - E\left[Y'|t=0\right]}{E\left[Y'|t=0\right]}\right) \tag{14}$$

To derive the estimated marketing effects in Table 6 we use the marginal effect estimates from estimating equation (11) and pre-to-post GDPR ratios of the usage metrics for each channel. Then, using the direct channel estimates, we can estimate θ by equation (13). Plugging θ into equation (14) yields estimates of δ' for each channel. Our resulting estimates

[24]We do not observe the real effect of the GDPR on non-consenting users.

Table 6: Marketing effect estimates: last-touch attribution sample

Channel	Direct traffic	Display	Email
		Pageviews	
Marginal recorded effect	-8.30%	-25.04%	-9.67%
Consent-adjusted marketing effect			
Marketing effect (δ', relative)	-	-20.01%	-3.77%
Marketing marginal effect (median firm weekly)	-	-650.04	-348.32
		Revenue	
Marginal recorded effect	-10.67%	-19.57%	-16.91%
Consent-adjusted marketing effect			
Marketing effect (δ', relative)	-	-20.25%	-6.56%
Marketing marginal effect (median firm weekly)	-	-$78.70	-$505.28

Note: The marketing marginal effect is computed relative to the weekly outcomes of the respective median last-touch attribution dashboard pre-GDPR.

of δ' are presented in Table 6. Table 6 also includes the corresponding level estimates for each δ' in pageviews and revenue, respectively. To construct level estimates for a representative firm, we multiply the marketing effect estimates by their channel share from Table 2 to estimate a cumulative marketing effect. Then, this cumulative marketing effect is multiplied by the median firm's weekly pre-GDPR outcomes. We focus on the display and email channel estimates as these are most directly attributable to marketing.[25]

For both pageviews and revenue we estimate that the display and email channels are negatively impacted by the GDPR. For pageviews, across both channels, we estimate a real decrease of 998 pageviews per week for the median firm. This represents a real decline of 0.4%, which is 9.4% of our last-touch attribution aggregate point estimate of 10,635 pageviews. Most of this decline is due to the display channel. For the e-commerce sample, we estimate a real loss of $584 per week for the median firm. This represents a 0.6% drop in real revenue, or 7.6% of our last-touch attribution aggregate estimate of -$7,686 dollar per week. Note that our channel shares in Table 2 are not conditional on using either the display or email channels, and thus our level estimates likely understate the true effect for firms that rely on these channels.[26] These small declines in real traffic resemble the 4% decline in traffic to EU publishers measured by Lefrere et al. (2020).

[25]The search and social channels capture both paid and organic traffic to the site. These channels also rely on first-party rather than third-party data, so search in particular should be less affected by the GDPR. The expected sign of any effects is therefore ambiguous.

[26]Alternatively, we can construct our level estimates using last-touch attribution shares calculated conditional on the firm reporting that channel (see our discussion of Table 2 in Section 3.4). Using conditional shares leads to slightly larger estimates, with weekly estimated decreases of 2,517 pageviews and $743 dollars. In percentage terms, our conditional-share marketing effects account for 15.6% of our pageviews point estimate and 9.3% of our revenue estimate.

These marketing effect estimates are conservative. We rely on last-touch attribution measures of marketing effectiveness which ignore the cumulative effect of marketing over time. Furthermore, these measures ignore cross-channel spillovers in general and the spill-overs from display and email advertising to direct traffic in particular: i.e., the causal effect of ads on direct navigation to the site. Budak et al. (2016) also rely on referral link data in their analysis of e-commerce site marketing, which shares these same limitations. Budak et al. (2016) report similar last-touch attribution shares for display (3%) and email advertising (7%). Finally, these estimates only capture the marketing effect on consenting users. Therefore, we understate the marketing effect by not accounting for any change in the real outcomes of non-consenting users.

7 Conclusion

Online firms leverage personal data to analyze consumer behavior, market themselves, and generate revenue. Privacy regulation limits personal data use with important consequences for online firms. We study the economic consequences of the GPDR for a large and diverse collection of online firms. Our Adobe Analytics data reveal the site performance of 1,084 firm dashboards. Relative to the previous year, we show that recorded pageviews fall by 11.7% and e-commerce recorded revenue falls 13.3% from EU users after the GDPR. However, these data alone do not distinguish between the real and recording effects on the GDPR. We propose a model to separate the GDPR's real effect on the volume of site visits and the GDPR's consent effect on the recording of site visit outcomes. We conclude that consent accounts for at least 7% of the recorded pageview estimate and at least 29% of the recorded revenue estimate. We also provide conservative estimates for the contribution of GDPR's real effect on personalized marketing. The marketing effect alone represents 9.4% of the recorded pageview estimate and 7.6% of the recorded revenue estimate.

We acknowledge important limitations to our research. First, while our data captures a significant portion of the online economy, the firms that use Adobe Analytics are selected—especially those that also track last-touch attribution data. Second, the GDPR's global repercussions make selecting a control group challenging, though we show our findings are robust to alternative control group choices. Third, we cannot separate the consent and real effects of the GDPR without making assumptions on how they differentially affect recorded site analytics data. Future research in this area could try to identify website compliance approaches or examine user panel data. Fourth, we analyze the economic consequences for firms, but do not provide a full welfare analysis of the GDPR. This is a general challenge in privacy settings due to the "privacy paradox" (Athey et al., 2017; Barnes, 2006) whereby

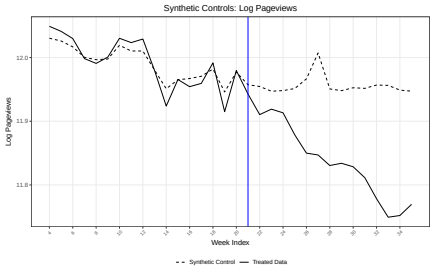
individual stated and revealed preferences for privacy diverge. Here, we lack consumer-level decision data to identify the demand for privacy. Fifth, we do not undertake a full profit analysis in this paper as we do not observe costs. Finally, we evaluate the early impact of the GDPR as it was interpreted by online firms in 2018. This period featured both limited regulatory enforcement and website compliance efforts that EU regulators broadly considered to be inadequate (Data Protection Commission, 2020b; Autoriteit Persoonsgegevens, 2019). Our results reflect this reality.

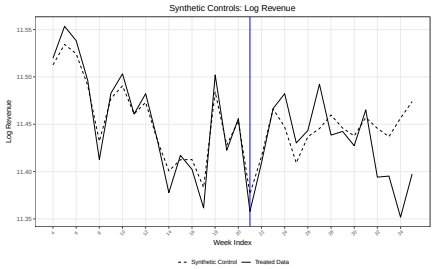
Our study offers several takeaways for privacy regulators. Data minimization is a key principle of the GDPR. EU regulators believe that the collection of web analytics data can pose a privacy risk to EU users and EU regulators therefore want websites to obtain GDPR-compliant consent from users. We document modest progress towards this goal and bound the average non-consent rate to be between [4.1%, 15.4%] for our full sample and [5.2%, 17.7%] for our e-commerce sample. Under the assumption that direct traffic to websites is affected by consent exclusively, we can point identify the average non-consent rate to be 8.5% for all dashboards and 15.2% for e-commerce dashboards (in our last-touch attribution samples). These results are consistent with the prevailing industry practice in 2018 whereby websites applied a *de facto* opt-out approach to consent. Despite this and concerns of consent fatigue (European Data Protection Board, 2020), a substantial minority of EU users apparently make the effort to register their non-consent preferences. We provide evidence from our e-commerce dashboards that smaller firms obtain lower consent rates, which suggests the GDPR may have consequence for competition. We only observe 3 of 1,084 dashboards that appear to adopt a stricter consent standard sought by regulators: we classify these dashboards as such because they exhibit over 50% reductions in recorded traffic and increases in recorded bounce rates of over 20%. These dashboards may foretell a future where regulators enforce GDPR-compliant opt-in consent.

Despite evidence of incomplete compliance, our results illuminate real consequences of the GDPR for online firms. First, we show larger drops in pageviews and revenue from users who click on a display ad or a marketing email—channels that rely on personal data. After adjusting for consent, we provide conservative estimates for the GDPR's real effect on outcomes generated via the display and email marketing channels: for the median firm, we estimate a real decrease of 0.4% in weekly pageviews in the full sample and 0.6% in weekly revenue in the e-commerce sample. Second, sites that show ads will lose personalized ad revenue from users who do not consent to data processing. Our back-of-the-envelope calculations put these costs between 0.4% and 6.1% in ad revenue. Third, limiting analytics data may hamper online firm's ability to create value from this data. This concern is limited by low non-consent rates on average but could become acute under stricter enforcement.

Finally, we examine the concern from firms that GDPR consent menus create friction that reduces user browsing. At least in 2018, we do not find evidence to support this concern over privacy frictions.

Figure 5: Synthetic controls results: fitted trends





Appendices

For online publication only.

A Robustness

A.1 Synthetic controls

Section 4.2.2 discusses the motivation for implementing synthetic controls. Here we discuss the results and methodology behind this approach in more detail.

Figure 5 plots average outcomes for our treated group with solid line and our fitted control group in a dotted line. The vertical line marks the implementation of the GDPR. The details of constructing our control group can be found in Section A.1.1. For both pageviews and revenue, Figure 5 illustrates a well fitted control group in the pre-GDPR period. For pageviews, the predicted control group deviates from our 2018 outcomes in the post-GDPR period - congruent with Figure 2. In contrast, our predict control group continues to match the treated group for revenue well into the post-GDPR period. Our point estimates for each outcome are presented in Table 4. The synthetic controls method estimates a treatment effect of 8.7% for pageviews and 1.4% for revenue.

Constructing standard errors in a synthetic controls setting with an elastic net is not a well understood problem. To provide some sense of the robustness of our synthetic controls results, we appeal to Abadie et al. (2010). Abadie et al. (2010) implement placebo studies by asking "how often would we obtain results of this magnitude if we had chosen a random (counterfactual) treated unit, rather than the factual treated unit?" In our context, we select a placebo treated unit from our control unit, fit a control group using our synthetic controls procedure, and then estimate a placebo treatment effect. We detail this procedure in Section A.1.2. In comparing the placebo and factual synthetic controls, we follow Abadie et al. (2010) in using an adjusted mean squared prediction error statistic (AMSPE).

Figure 6: Synthetic control results: placebo tests

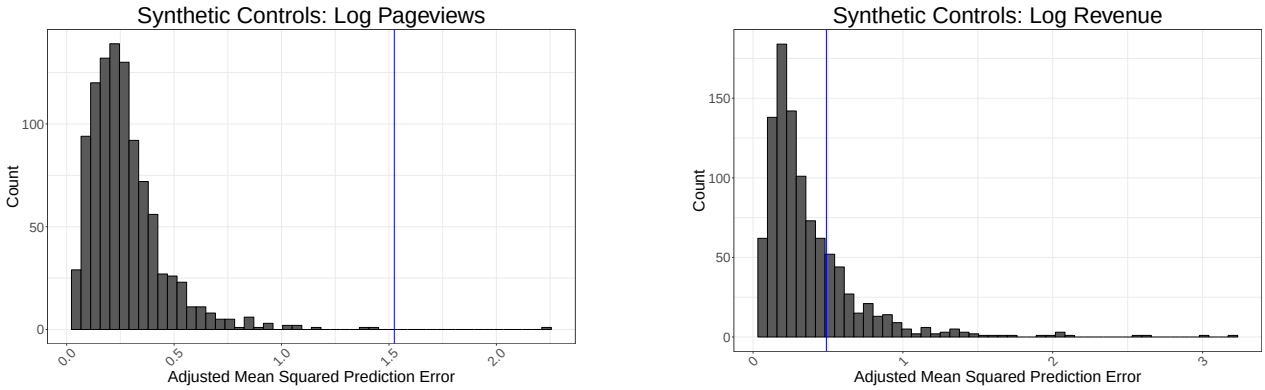


Figure 6 presents histograms of our 1000 placebo studies for pageviews and revenue. The histograms reflect the distribution of AMSPE estimates across our 1000 placebo trials, for each outcome. The vertical line marks the AMSPE of our true estimated treatment effect, 1.5 for pageviews and 0.49 for revenue. For pageviews, our true AMPSE is at the 99.8th quantile of the placebo AMPSE distribution, which suggests that recovering a treatment effect of the magnitude presented in Table 4 column (4) by chance is highly unlikely. The AMPSE for revenue is at the 78.4th quantile of the placebo AMPSE distribution. Our synthetic controls estimate for revenue is therefore less precise than our pageviews measure, and somewhat more likely to occur by chance.

A.1.1 Synthetic controls: cross-validation & estimation

For the following discussion, T_0 will be the period before which the intervention takes place, $Y(0)$ is the counterfactual outcome, and $Y(1)$ is the observed outcome of the treatment. In our setting, we will use 2017 dashboard log outcome data as control units and the average (across dashboards) of 2018 log outcome data as our treatment unit. That is, for each outcome variable, we have a treatment unit and a set of control units:

$$Y_t = \frac{1}{N}\sum_{i}^{N} \log\left(y_{it}^{2018} + 1\right) \tag{15}$$

$$C = \{C_{it} = \log\left(y_{it}^{2017} + 1\right) \forall\ i\} \tag{16}$$

We estimate weights such that the weighted combination of C_{it} best matches Y_t. In our setting, we have 17 pre-treatment time periods and 1084 control units, or $N >> T_0$. We follow Doudchenko and Imbens (2016) in using an elastic net to construct our control group. See Zou and Hastie (2005) for a detailed discussion of elastic nets and their properties. In

brief, we fit a model with the following objective function:

$$Q(\mu, \omega | Y_t, C_{it}; \alpha, \lambda \text{ for } t < T_0) = ||Y_t - \mu - \omega C_{it}||_2^2 + \lambda \cdot \left(\frac{1-\alpha}{2} ||\omega||_2^2 + \alpha ||\omega||_1 \right) \tag{17}$$

Where μ is a constant, ω is a vector of length N of weights, and α and λ are penalty parameters chosen by the econometrician.

We choose penalty parameters using a modified version of the cross validation routine proposed in Doudchenko and Imbens (2016). In particular, for a proposed pair of penalty parameters, $\{\alpha', \lambda'\}$, we construct pseudo treated units as follows. First, we partition C into B random partitions of size b. We will refer to a partition as C^b. Each C^b is used to construct a pseudo treated unit, $Y_t^{C^b}$, by taking the average over units $i \in C^b$. We use $\tilde{C} = C \backslash C^b$ as the control units for pseudo treated unit $Y_t^{C^b}$. An elastic net is fitted, using only *pre-intervention* data, to obtain $\{\hat{\mu}^b, \hat{\omega}^b\}$. That is: $\{\hat{\mu}^b, \hat{\omega}^b\} = \text{argmin}_{\mu,\omega} \sum_{t=1}^{T_0} \left(Y_t^{C^b} - \mu - \omega \tilde{C}_{it} \right)^2 + \lambda' \cdot \left(\frac{1-\alpha'}{2} ||\omega||_2^2 + \alpha' ||\omega||_1 \right)$. Given the weights estimated above and using the proposed penalty parameters $\{\alpha', \lambda'\}$, we predict the outcome for $Y_t^C(0)$ in $t > T_0$ and construct the mean squared error for each B.

$$Y_t^{C^b}(0) = \hat{\mu}^b + \hat{\omega}^b \tilde{C}_{it} \tag{18}$$

$$CV_B(\alpha', \lambda') = \frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(Y_t^{C^b}(1) - Y_t^{C^b}(0) \right)^2 \tag{19}$$

Model performance is then evaluated using the average of the cross validated mean squared error across our B partitions:

$$CV(\alpha', \lambda') = \frac{1}{B} \sum_b CV_b(\alpha', \lambda') \tag{20}$$

Finally, tuning parameters are chosen such that $\{\alpha, \lambda\} = argmin_{\alpha', \lambda'} CV(\alpha', \lambda')$. Using these tuning parameters, we recover the vector of weights ω needed to construct our synthetic control.

We search over a grid of $\alpha \in [.01, .99]$ in increments of .01 and take advantage of the λ validation built into the glmnet.R package (Friedman et al., 2010). For each $\{\alpha', \lambda'\}$ we partition the control units into $B = 10$ samples - analogous to 10 cross-fold validation. We repeat the above procedure 100 times for each outcome variables to construct a set of ω vectors, which we average to construct our final weights. We then generate our estimates of the treatment effect by differencing the average levels of the treated unit and the synthetic control in the post-GDPR period.

A.1.2 Synthetic controls: Placebo routine

We can appeal to Abadie et al. (2010) to get a sense of how reasonable our results are. In particular, the following exercise asks: "how large would our prediction error be had treatment not occurred?" To construct this counterfactual, our procedure is as follows:

- Randomly sample $n = 10$ units from C
- Construct our pseudo-treated unit as $C_t^{psuedo} = \frac{1}{n} \sum_{i \in sample} C_{it}$
- Fit an elastic net to C_t^{psuedo} as described in appendix A.1.1
- Calculate the adjusted mean squared prediction error (Abadie et al., 2010):

$$\frac{T_0}{T - T_0} \frac{\sum_{t=T_0}^{T} \left(Y_t^{psuedo}(1) - Y_t^{C(psuedo)}(0) \right)^2}{\sum_{t=0}^{T_0} \left(Y_t^{psuedo}(1) - Y_t^{C(psuedo)}(0) \right)^2} \tag{21}$$

  - That is, we calculate the mean squared prediction error and scale it by the mean squared fitting error

- Repeat 1000 times for each outcome

The results of this procedure are presented in Figure 6.

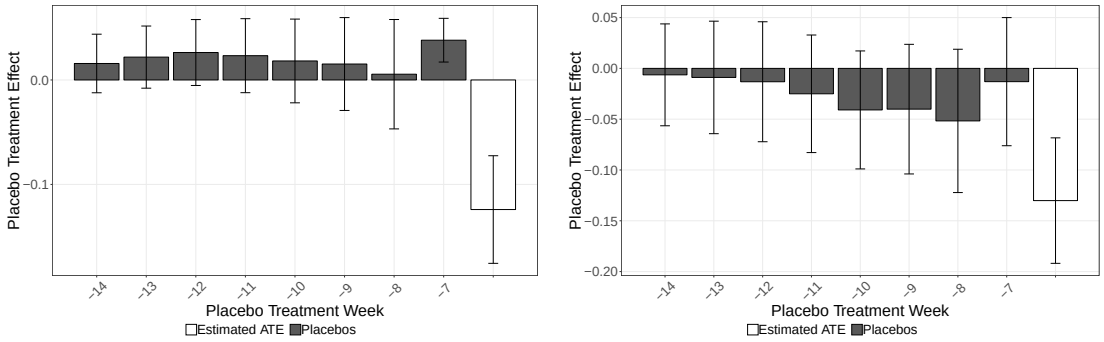
B Placebo tests

We can run placebo tests to assess the potential for false positive effect estimates. We implement placebo tests by first choosing a counterfactual treatment week from the pre-GDPR period of our data. Then, equation (22) is estimated using data from before April 25th (pre-GDPR). We exclude one-month before the implementation of the GDPR in order to omit any anticipatory behavior. This procedure is repeated for placebo treatment dates ranging from 14 to 7 weeks prior to May 25th, for a total of 8 placebo tests. These placebo dates are chosen in order to provide adequate pre-trends and post-trends in the data (at least 3 data points before and after the placebo treatment).

$$\log(y_{itw} + 1) = \alpha \mathbb{1}\{2018\}_t + \beta_p \left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post Placebo}\}_w \right) + \theta_i + \eta_w + \epsilon_{itw} \tag{22}$$

The primary coefficient of interest is β_p. Fixed effects are included as in equation (2) and all standard errors are clustered at the dashboard-week level. Significant point estimates are

Figure 7: Main effect panel difference pre-trend placebo tests



indicative of false positives; a prevalence of false positives may undermine the credibility of our point estimates.

Our placebo results are presented in Figure 7 in grey. Figure 7 also includes our estimated treatment effect in white, to aid in comparison. The placebo results demonstrate the robustness of our identification strategy and point estimates in Table 3. No estimates match the magnitude of our main results, nor are any of the placebo tests as statistically significant.

C Marginal effects

The models in Sections 4 and 6.3 are non-linear models and therefore rely on marginal effects for interpretation. In this section we detail the construction of these marginal effects. Our models are generally of the form:

$$log(y_{itw}+1) = \alpha \mathbb{1}\{2018\}_t + \beta\left(\mathbb{1}\{2018\}_t \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w\right) + \eta_i + \xi_w + \epsilon_{itw}$$

where β captures the GDPR effect. First, we estimate the above regression using the data. Then, for the post-GDPR period, we use our estimates to construct predicted outcomes for both the treated and a counterfactual untreated group:

$$y_{itw}^{untreated} = \exp\left(\alpha + \eta_i + \xi_w + \frac{\sigma^2}{2}\right) - 1$$
$$y_{itw}^{treated} = \exp\left(\alpha + \beta + \eta_i + \xi_w + \frac{\sigma^2}{2}\right) - 1$$

Where we have included variances ($Var[\epsilon_{itw}] = \sigma^2$) to account for the expected value of the (log-normal) error terms. These predictions are at the dashboard-week level. We construct

the marginal effects using the subsample of dashboard-weeks from 2018 after GDPR enforcement; we denote the set of post-GDPR weeks by W_{post}. We then compute the average marginal effect (AME) as follows:

$$AME = \frac{1}{N}\frac{1}{|W_{post}|}\sum_{i=1}^{N}\sum_{w\in W_{post}}\frac{y_{iw}^{treated} - y_{iw}^{untreated}}{y_{iw}^{untreated}} \tag{23}$$

where $|W_{post}|$ denotes the number of post-GDPR weeks in the sample and N is the number of dashboards in the sample.

D Heterogeneity results

D.1 Regulatory Enforcement Regressions

To construct our index of regulatory enforcement beliefs, we leverage a European Commissions survey from 2008 of 4,835 data controllers across all countries in the EU European Commission (2008). The survey asks to what extent data controllers agree or disagree with "the data protection law in (OUR COUNTRY) is interpreted and applied more rigorously than in other Member States." Responses are recorded on a four point scale and include a no-response option. We exclude the non-responses and construct a response-weighted average index for each country in the EU that takes values from 0 (all responses are "totally disagree") to 1 (all responses are "totally agree"). We then standardize the index to have mean of zero and variance of one. We additionally utilize data on GDP per capita for each EU country in 2018 from the World Bank, which we expect correlates with advertising and e-commerce revenue.[27]

We explore the role of regulatory strictness empirically, by interacting our panel differences estimator with our regulatory strictness measure. We estimate the following equation:

$$\begin{aligned}\log(y_{itw}+1) = {} & \alpha_1 \mathbb{1}\{2018\} + \alpha_2 \mathbb{1}\{\text{Index x } \mathbb{1}\{2018\} + \alpha_3 \mathbb{1}\{\text{GDP}\} \text{ x } \mathbb{1}\{2018\} + \\ & \gamma_1 \mathbb{1}\{\text{Index}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w + \gamma_2 \mathbb{1}\{\text{GDP}\}\text{x}\mathbb{1}\{\text{Post GDPR}\}_w + \\ & \beta \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\} + \\ & \beta_{GDP} \mathbb{1}\{\text{GDP}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\} + \\ & \beta_{index} \mathbb{1}\{\text{Index}\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}_w \text{ x } \mathbb{1}\{2018\} + \eta_i + \xi_w + \epsilon_{itw}\end{aligned} \tag{24}$$

Our interaction coefficient of interest is β_{index}. Following Jia et al. (2018), we also include

[27] https://Data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=PL-GR-PT-DE-EU

Table 7: Regulator Heterogeneity Regressions

	(1)	(2)	(3)	(4)
Dependent Variable	log(Pageviews + 1)		log(Revenue + 1)	
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \text{Index}$	-0.056	-0.040	-0.047	-0.041
	(0.006)	(0.006)	(0.016)	(0.019)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	-0.074	0.041	-0.096	-0.040
	(0.022)	(0.029)	(0.022)	(0.075)
$\mathbb{1}\{2018\} \times \text{Index}$	-0.007	-0.008	0.030	0.008
	(0.005)	(0.004)	(0.007)	(0.009)
$\mathbb{1}\{\text{Post GDPR}\} \times \text{Index}$	0.032	0.027	0.002	-0.014
	(0.007)	(0.009)	(0.010)	(0.012)
$\mathbb{1}\{2018\}$	0.056	0.049	0.163	-0.025
	(0.008)	(0.024)	(0.015)	(0.059)
$\mathbb{1}\{2018\} \times \text{GDP}$		0.019		0.469
		(0.056)		(0.151)
$\mathbb{1}\{\text{Post GDPR}\} \times \text{GDP}$		0.101		0.345
		(0.054)		(0.128)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \text{GDP}$		-0.294		-0.138
		(0.079)		(0.186)
RSID + Week FE	Y	Y	Y	Y
R^2	0.969	0.969	0.962	0.962
N	68,960	68,960	22,436	22,436

*p<0.1; **p<0.05; ***p<0.01

Note: Standard errors clustered at the Dashboard + Week level

an interaction with country-level GDP per capita, because income is a potential confound that correlates with regulatory strictness. Results of this regression are presented below.

D.2 Size heterogeneity results

Here we present results examining heterogeneity in GDPR treatment effect estimates as a function of site size. A key point of policy interest is the competitive ramifications of the GDPR. Other work has demonstrated that the GDPR lead to greater market concentration in the website technology services sector (Johnson et al., 2020; Peukert et al., 2020). Larger sites may benefit from the GDPR as they may more easily obtain consent (Campbell et al., 2015) and may benefit from economies of economies of scale in compliance. We look for heterogeneous effects of the GDPR across dashboard size by interacting our main specification (equation 2) with an indicator for large dashboards. This indicator equals one for dashboards with above-median pageviews in the pre-GDPR period.

Table 8 presents the results of these regressions The coefficient on $\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$ measures the impact of the GDPR on small dashboards. In column (1), our point estimate

Table 8: Treatment effect heterogeneity by dashboard size

Dependent Variable	(1) log(Pageviews + 1)	(2) log(Revenue + 1)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Large Dashboard}\}$	0.040	0.099
	(0.034)	(0.049)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	-0.144	-0.192
	(0.035)	(0.039)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Large Dashboard}\}$	-0.057	-0.060
	(0.024)	(0.045)
$\mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Large Dashboard}\}$	-0.039*	-0.055
	(0.020)	(0.044)
$\mathbb{1}\{2018\}$	0.081	0.223
	(0.018)	(0.042)
Average Marginal Effect		
Small Dashboard	-13.43%	-17.43%
Large Dashboard	-9.87%	-8.87%
RSID + Week FE + Size FE	Y	Y
R^2	0.969	0.962
N	69,344	22,436

Note: Standard errors clustered at the Dashboard + Week level, *p<0.1; **p<0.05; ***p<0.01

for recorded pageviews is -0.14, which is close to our main effect point estimate. We estimate a statistically insignificant interaction coefficient of 0.04 for large dashboards, indicating that large dashboards experience a decline in pageviews due to the GDPR that is indistinguishable from small dashboards. In contrast, our recorded revenue marginal effect estimates in column (2) imply small e-commerce firms may be suffering more than large firms, with an estimated decline of -17.4% for small firms and -8.9% for large firms. As 17% of EU enterprises derive some revenue from online sales (EuroStat, 2020), the disparate effect on small firms is troubling if it reflects real outcomes. A remaining question is what drives this heterogeneity. We explore this question in Sections 6 and 6.2 and provide evidence that heterogeneity in consent rates is driving this result.

D.2.1 Size Heterogeneity Site Usage Results

Table 9 presents the full regression results examining heterogeneity in site usage.

E Last-touch attribution regression estimates

Table 10 presents the coefficient estimates for the different last-touch attribution channels. Note that these estimates are graphed in Figure 4.

Table 9: Treatment effect heterogeneity by site size: usage metrics

Dependent Variable	(1) Pageviews per visit	(2) Revenue per visit
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Large Dashboard}\}$	-0.050	-0.310
	(0.131)	(0.189)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\}$	0.225	0.327
	(0.113)	(0.163)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Large Dashboard}\}$	0.176	0.368*
	(0.125)	(0.188)
$\mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Large Dashboard}\}$	0.132	0.425
	(0.113)	(0.182)
$\mathbb{1}\{2018\}$	-0.379	-0.145
	(0.114)	(0.169)
Implied θ bounds		
Small Dashboard	[4.16%, 17.56%]	[6.91%, 25.54%]
Large Dashboard	[2.99%, 13.04%]	[2.27%, 9.28%]
RSID + Week FE + Size FE	Y	Y
R^2	0.725	0.759
N	69,344	22,436

*p<0.1; **p<0.05; ***p<0.01

Note: Standard errors clustered at the Dashboard + Week level

Table 10: Last-touch attribution channel ATE regression estimates

Dependent Variable	(1) Pageviews per visit	(2) Revenue per visit
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Direct Traffic}\}$	-0.087	-0.113
	(0.029)	(0.042)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Search}\}$	-0.077	-0.033
	(0.028)	(0.039)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Display}\}$	-0.287	-0.218
	(0.076)	(0.106)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Email}\}$	-0.101	-0.185
	(0.042)	(0.070)
$\mathbb{1}\{2018\} \times \mathbb{1}\{\text{Post GDPR}\} \times \mathbb{1}\{\text{Social}\}$	-0.240	-0.034
	(0.040)	(0.064)
RSID + Week FE	Y	Y
R^2	0.950	0.940
N	137,469	40,373

*p<0.1; **p<0.05; ***p<0.01

Note: Standard errors clustered at the Dashboard + Week level

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Adjerid, I., A. Acquisti, R. Telang, R. Padman, and J. Adler-Milstein (2016). The impact of privacy regulation and technology incentives: The case of health information exchanges. *Management Science 62*(4), 1042–1063.

Adobe (2019). Adobe analytics documentation.

Adobe (2020, March). Adobe unveils first digital economy index.

Aridor, G., Y.-K. Che, W. Nelson, and T. Salz (2020). The economic consequences of data privacy regulation: Empirical evidence from gdpr. *Available at SSRN*.

Article 29 Data Protection Working Party (2012, June). Opinion 04/2012 on cookie consent exemption. Technical report, Article 29 Data Protection Working Party.

Athey, S., C. Catalini, and C. Tucker (2017). The digital privacy paradox: Small money, small costs, small talk.

Autoriteit Persoonsgegevens (2019, December). Ap: veel websites vragen op onjuiste wijze toestemming voor plaatsen tracking cookies.

Barnes, S. B. (2006). A privacy paradox: Social networking in the united states. *First Monday 11*(9).

Berman, R. and A. Israeli (2020). The added value of descriptive analytics: Evidence from online retailers. HBS Working Paper.

Bresnahan, T. F., E. Brynjolfsson, and L. M. Hitt (2002, February). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence*. *The Quarterly Journal of Economics 117*(1), 339–376.

Brynjolfsson, E., L. M. Hitt, and H. H. Kim (2011). Strength in numbers: How does data-driven decision-making affect firm performance? *SSRN eLibrary*.

Brynjolfsson, E. and J. Oh (2012). The attention economy: Measuring the value of free digital services on the internet. In *ICIS*.

Budak, C., S. Goel, J. Rao, and G. Zervas (2016). Understanding emerging threats to online advertising. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 561–578. ACM.

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

Cavallo, A. and R. Rigobon (2016, May). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives 30*(2), 151–78.

Commission Nationale de l'Informatique et des Libertés (2020a, October). Cookies et autres traceurs : la cnil publie des lignes directrices modificatives et sa recommandation.

Commission Nationale de l'Informatique et des Libertés (2020b, September). Délibération 2020-091 du 17 septembre 2020. Technical report, Commission Nationale de l'Informatique et des Libertés.

Competition and Markets Authority (2020). Appendix f: the role of data in digital advertising. In *Online platforms and digital advertising market study*.

Data Protection Commission (2020a, April). Guidance note: Cookies and other tracking technologies. Technical report, Data Protection Commission.

Data Protection Commission (2020b, April). Report by the data protection commission on the use of cookies and other tracking technologies. Technical report, Data Protection Commission.

de Matos, M. G. and I. Adjerid (2019). Consumer behavior and firm targeting after GDPR: The case of a telecom provider in Europe. Working paper.

Dolfen, P., L. Einav, P. J. Klenow, B. Klopack, J. D. Levin, L. Levin, and W. Best (2019, February). Assessing the gains from e-commerce. Working Paper 25610, National Bureau of Economic Research.

Doudchenko, N. and G. W. Imbens (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Working Paper 22791, National Bureau of Economic Research.

European Commission (2008, January). Flash eurobarometer 226: Data protection in the european union : Data controllers' perceptions. https://data.europa.eu.

European Data Protection Board (2020, May). Guidelines 05/2020 on consent under regulation 2016/679. Technical report, European Data Protection Board.

EuroStat (2020, April). Online sales continue to grow for eu enterprises.

Forrester (2017, December). The forrester wave: Web analytics, q4 2017. Online Report.

Friedman, J. H., T. Hastie, and R. Tibshirani (2010). Glmnet: Lasso and elastic-net regularized generalized linear models. Technical report, http://CRAN.R-project.org/package=glmnet.

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

Goolsbee, A. D. and P. J. Klenow (2018). Internet rising, prices falling: Measuring inflation in a world of e-commerce. In *AEA Papers and Proceedings*, Volume 108, pp. 488–92.

Information Commissioners Office (2019). Cookies – what does 'good' look like?

Information Commissioner's Office (2019, October). Disclosure irq0873632.

Jia, J., G. Z. Jin, and L. Wagman (2018). The short-run effects of GDPR on technology venture investment. Technical report, National Bureau of Economic Research.

Jia, J., G. Z. Jin, and L. Wagman (2019). GDPR and the home bias of venture investment.

Johnson, G., S. Shriver, and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science 39*(1), 33–51.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2020). Privacy & market concentration: Intended & unintended consequences of the gdpr.

Joskow, P. L. and N. L. Rose (1989). Chapter 25 the effects of economic regulation. Volume 2 of *Handbook of Industrial Organization*, pp. 1449–1506. Elsevier.

Ke, T. T. and K. Sudhir (2020). Privacy rights and data security: Gdpr and personal data driven markets. *Available at SSRN 3643979*.

Laffont, J.-J. (1994). The new economics of regulation ten years after. *Econometrica 62*(3), 507–537.

Lambrecht, A., K. Seim, and C. Tucker (2011). Stuck in the adoption funnel: The effect of interruptions in the adoption process on usage. *Marketing Science 30*(2), 355–367.

Lefrere, V., L. Warberg, C. Cheyre, V. Marotta, and A. Acquisti (2020). The impact of the gdpr on content providers.

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

Long, L. (2020, January). How to a/b test to optimize the data collection consent experience for users.

Manski, C. (2007). *Identification for prediction and decision*. Harvard Univ Pr.

Manski, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics 5*(1), 145–174.

Manski, C. F. and J. V. Pepper (2018, 05). How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions. *The Review of Economics and Statistics 100*(2), 232–244.

Matte, C., C. Santos, and N. Bielova (2020, October). Purposes in IAB Europe's TCF: which legal basis and how are they used by advertisers? In *APF 2020 - Annual Privacy Forum*, Lisbon, Portugal, pp. 1–24.

McDonald, A. M., R. W. Reeder, P. G. Kelley, and L. F. Cranor (2009). A comparative study of online privacy policies and formats. In I. Goldberg and M. J. Atallah (Eds.), *Privacy Enhancing Technologies*, Berlin, Heidelberg, pp. 37–55. Springer Berlin Heidelberg.

Miller, A. R. and C. Tucker (2009). Privacy protection and technology diffusion: The case of electronic medical records. *Management Science 55*(7), 1077–1093.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2020). European privacy law and global markets for data.

PricewaterhouseCoopers (2018). Pulse survey: GDPR budgets top $10 million for 40% of surveyed companies. https://www.pwc.com/us/en/services/consulting/library/general-data-protection-regulation-gdpr-budgets.html.

Quantcast (2018, July). Quantcast choice powers one billion consumer consent choices in two months since gdpr. Press Release.

Ravichandran, D. and N. Korula (2019, August). Effect of disabling third-party cookies on publisher revenue. Technical report, Google Inc.

Sanchez-Rola, I., M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Verviker, and I. Santos (2019). Can I opt out yet? GDPR and the global illusion of cookie control. In *ACM ASIACCS 2019*.

Snelders, E., L. Worp, and S. Song (2020). A future without advertising cookies? it's possible! Technical report, Ster.

Sørensen, J. and S. Kosta (2019). Before and after GDPR: The changes in third party presence at public and private european websites. In *The World Wide Web Conference*, WWW '19, New York, NY, USA, pp. 1590–1600. ACM.

TrustArc (2018, July). Gdpr compliance status: A comparison of us, uk and eu companies. Technical report, TrustArc.

U.S. Census Bureau (2019). 2017 e-stats report: Measuring the electronic economy.

Zhuo, R., B. Huffaker, K. Claffy, and S. Greenstein (2019, November). The impact of the general data protection regulation on internet interconnection. Working Paper 26481, National Bureau of Economic Research.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology) 67*(2), 301–320.