Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes

Samuel Goldberg[*] Garrett Johnson[†] Scott Shriver[‡]

July 16, 2019

Abstract

In May 2018, the European Union began enforcing the General Data Protection Regulation (GDPR), which endowed EU citizens with new personal data rights and imposed new responsibilities on firms. Privacy regulation increases the firm's cost of collecting consumer data which makes matching with users more costly. As such, the GDPR has the potential to reduce both the amount of traffic to a website as well as the amount and quality of web outcome data stored for analytics purposes. We examine the impact of the GDPR on European web traffic and E-commerce sales using web analytics data from a diverse set of 1508 firms that use the Adobe Analytics platform. Using a difference-in-differences approach, we show that recorded pageviews and recorded revenues fall by about 10% for EU users after the GDPR's enforcement deadline. The extensive margin drives these changes as a user's average time on site and average page views per visit stay constant.

1 Introduction

The European Union (EU) is a global leader in privacy regulation whose constitution enshrines the right to privacy. In May 2018, the EU began enforcing its General Data Protection Regulation

[*] Kellogg School of Management, Northwestern University. Samuel.Goldberg@kellogg.northwestern.edu

[†] Questrom School of Business, Boston University. garjoh@bu.edu

[‡] Leeds School of Business, University of Colorado at Boulder. Scott.Shriver@colorado.edu

The authors would like to thank Adobe for providing the data, Taylor Schreiner for his support within Adobe, Brett Gordon as well as seminar participants at Kellogg School of Management, INSEAD, Lehigh University, Marketing Science and Boston University.

(GDPR), a landmark privacy law that defines individual privacy rights and restricts how firms can use personal data. By protecting individual privacy, the GDPR can hurt firms that rely on customer analytics to make decisions and for personalized marketing. In particular, online firms collect detailed web analytics data on how users navigate through—and arrive at—websites using platforms like Adobe Analytics. Online firms use web analytics data to better draw users to their sites and to improve site content and usability. Under the GDPR, firms may choose to collect less web analytics data or may find that fewer users consent to data collection. The GDPR also increases the cost of personalized marketing channels like e-mail and display ads that draw users to websites. The GDPR may even change user preferences for browsing online by making privacy more salient and actionable. Thus, the GDPR could hurt online firms: 1) directly, by restricting online advertising and changing user browsing preferences, and 2) indirectly, by reducing the web analytics data that informs the firm's decisions.

We empirically investigate for the first time the impact of privacy policy on recorded web outcomes. We are among the first to study the GDPR, whose scale and scope has cost many firms millions of dollars in compliance costs (PWC 2018). Like economic studies of past privacy regulation (e.g. Goldfarb & Tucker 2011; Tucker & Miller 2009), we leverage the timing of regulatory enforcement as an event study. We further leverage proprietary data from Adobe Analytics to examine the impact of the GDPR on 1,500 online firms constituting over 1 billion weekly visits by EU residents. These firms include 128 of the top 1,000 global sites and feature a variety of content, E-commerce, and corporate sites. Using a difference-in-differences strategy, we find that recorded online outcomes fall about 10% across the board: page views, visits, orders, and revenue. This result is robust to another strategy combining frontier synthetic control and machine learning approaches (Doudchenko & Imbens 2016). Despite the total reduction, we see no change in user quality metrics, which suggests a firm-driven rather than a consumer-selection driven explanation.

The GDPR is a landmark privacy law that is inspiring a wave of privacy regulation in such countries as Brazil, India, Japan, and South Korea. Similarly, the United States is considering federal privacy regulation to harmonize state privacy laws led by California. The GDPR protects all personal data rather than just personally identifiable data or sensitive data categories like health data. The GDPR strengthens individual ownership rights over personal data by granting rights to access, correct, and delete personal data held by firms. Firms must minimize personal data processing and can only process personal data under limited and specific circumstances. One such circumstance is an individual's explicit opt-in consent, which is well suited to processing clickstream data. Requiring consent increases the cost of collecting web analytics data as well as the personalized digital marketing channels that depend on clickstream data.

Despite the GDPR's importance for the EU's 28 countries and beyond, little is known about the economic consequences of the law. We contribute the first study of the online economic impact of the GDPR. Contemporary economic research shows that the GDPR hurt venture capital investment (Jia, Jin, & Wagman 2018). Online, the GDPR led to a reduction in third party cookies (Libert, Graves, & Nielsen 2018) and updated online privacy statements (Degeling et al. 2019). Other authors studied the effectiveness of a GDPR consent campaign for obtaining consent and personal marketing (Godinho de Matos & Adjerid 2019). Past literature on privacy regulation shows that American health privacy laws slowed technology diffusion (Tucker & Miller 2011). Goldfarb & Tucker (2011) most resembles our own as they show the EU's 2002 e-Privacy Directive reduced online display ad effectiveness.

Our proprietary data from Adobe Marketing Cloud provides a broad view of the GDPR's online impact. Adobe's web analytics offering is the 4th most frequently installed analytics vendor in the category among top 10,000 sites.[2]. Ours is only the second study to use such Adobe data, after Goolsbee & Klenow (2018) who study online inflation. We see the web traffic of 1,500 firms from such diverse industries as media, travel, and retail across a mixture of content, corporate, and e-commerce sites. Crucially, our data differentiates users by location and arrival point so that we can identify EU users and the marketing channels that push them to sites. Our data include economically important outcomes for different site types: content sites monetize page views using advertising and E-commerce sites rely on online purchases. Our total data contain over 4.6 billion page views and $0.5 billion in revenue weekly from EU users.

We use the GDPR's May 2018 enforcement deadline as an event study. We use both difference-in-differences and synthetic control approaches to identify the impact of the GDPR. The former approach uses site activity from the EU in 2017 as a control in order to account for seasonal

[2] https://trends.builtwith.com/analytics/Adobe-Marketing-Cloud, accessed on June 5, 2019

differences. Across all sites, we estimate that recorded page views fall 9.7% and recorded site visits fall 9.9% post-GDPR. Among e-commerce sites, we estimate that recorded site outcomes fall 5.6% and recorded revenue falls 8.3%. For the median site, this corresponds to a $8000 weekly reduction in revenue. These reductions in recorded web outcomes are robust to a synthetic control approach based on Doudchenko & Imbens (2016) that uses machine learning to match the trend in pre-GDPR web outcomes to a combination of firms from 2017.

Our data offers clues on how the GDPR affects recorded web outcomes. Some firms stop sharing web analytics data with Adobe post-GDPR perhaps due to the GDPR's data minimization mandate. However, we eliminate these firms from our sample by construction: the drop in recorded web outcomes would otherwise be even larger. We also do not find evidence of user selection post-GDPR, whether due to user's changing the preference for sites post-GDPR or due to only recording data from consenting users. In particular, we see no change in average time-spent or page views per visit—common user quality metrics. In the future, we intend to use our data on how users arrive to the site (e.g. online ad or direct navigation) for evidence that the GDPR impacted site visits through its impact on advertising.

We proceed by overviewing the GDPR, then presenting the Adobe Analytics data and our results, before concluding.

2 The General Data Protection Regulation

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 as well as all customers of EU-based firms or firms with EU offices. The GDPR expands the definition of personal information beyond personally-identifiable data to include individual-level data like cookies and IP addresses. GDPR fines can reach the larger of 20 million euros or 4% of global turnover.

The GDPR accords new rights to individuals and responsibilities to data-processing firms. Individuals receive the right to access their personal data, correct data, erase data, and port data elsewhere as well as the rights to object to data processing and object to decisions based on automated processing. Under the GDPR, firms face both rights- and risk-related obligations. The rights-related obligations require that firms allow individuals to exercise their rights in an easy and timely manner. As for risk-related obligations, firms must appoint a Data Protection Officer to oversee compliance activities and must audit internal data processes. Also, firms must encrypt and anonymize personal data (data protection by design) as well as minimize data collection (data protection by default). In the event of a data breach, firms must promptly notify the regulator and affected individuals. These obligations impose potentially large compliance and opportunity costs on firms. Many firms are spending over 10 million dollars annually to comply with the law and many still are coming into compliance after May 25, 2018 (PWC 2018).

The GDPR defines the legal bases for processing personal data. Firms can process data in order to fulfill a contract or legal obligation, to protect the public interest and to protect the vital interest of the individual. Otherwise, firms may obtain an individual's consent. 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. Finally, firms can claim their own "legitimate interest" as a basis for data processing, though the GDPR cautions this can not override individual data rights. The GDPR thus increases the marginal cost of collecting and using individual data particularly when collecting consent. Below, we explain how the GDPR applies to, and may affect, web analytics.

2.1 Web analytics under the GDPR

The EU's current guidance emphasizes consent as the primary legal basis for using web analytics under the GDPR. Data processing is also allowed under a contractual obligation. For example, if a user initiates a purchase on an E-commerce site, that site can process an individual's name, address, and credit card information, under the contractual obligation clause, to complete the transaction. However, the EU's draft guidelines states that processing data for web analytics—even for the purpose of improving a service by the site—cannot rely on contractual obligation as web analytics are not necessary to fulfilling a contract. Instead, regulators have indicated that firms should seek consent or may use legitimate interest as a legal basis (EDPB 2019). Using legitimate interest is potentially legally risky as it is unclear whether web analytics would pass the EU's proscribed balance of interest test (Article 29 Working Party 2014). Prior to the GDPR, online firms collected

	Total Outcomes	**Share Recorded**
Firm-Driven	Advertising	Data Minimization
User-Driven	Privacy Salience	Consent

Table 1: Detailing the privacy mechanism

this data and may have provided an e-Privacy Directive cookie notice to users and offered a user opt-out.

By potentially reducing the share of recorded web outcomes, the GDPR creates an identification challenge in our web analytics data. In particular, we only observe *recorded* outcome measures in our data, which is the product of two terms:

$$Recorded\ Outcomes = Total\ Outcomes \times Share\ Recorded$$

This creates an identification problem, as both the total outcomes and propensity to record outcomes are likely to be impacted by the GDPR. Further, we lack direct firm or consumer data that would separate the two components. Each component can be affected by firm-driven or user-driven causes. The matrix in figure 1 identifies the principle mechanism in each case, which we detail below.

1. **Total Outcomes**

   (a) **Advertising Effect**: The higher costs of using personal information can affect personalized marketing channels that drive online traffic. For instance, the GDPR increased the legal risk associated with e-mail and online display advertising as both rely on personal data in the form of cookies or the e-mail lists. As such, the quality and quantity of advertising through these channels may fall. Browsing data suggests that e-mail and display ads precede 7 and 3% respectively of visits to e-commerce sites (Budak et al. 2016). Similarly, the previous EU privacy legislation has been shown to reduce ad effectiveness by 65% (Goldfarb & Tucker 2011). Therefore this may be an important effect. We will look for direct evidence for this mechanism using our data on the marketing channel that precede site visits.

   (b) **Privacy Salience Effect**: Overall site traffic may change as users become aware of how their information is used. GDPR enforcement brought ubiquitous privacy notices on websites that serve EU users. By increasing the salience of privacy concerns, these notices may have changed user preferences for how much time users spend online and which sites they frequent. We look for evidence of this channel in user's average time spent and page views per session, because we expect self-selection in the residual user web traffic.

2. **Share Recorded**

   (a) **Data Minimization Effect**: The GDPR requires that firms minimize the data they collect about users. Some firms may elect to no longer collect web analytics data or to reduce the duplication in the number of web analytics vendors. We avoid this explanation by construction: our data sample avoids firms who turn off web analytics at our data provider. However, we will struggle to detect firms that reduce data sharing but still report data for reasons unrelated to consumer consent.

   (b) **Consent**: User consent is the primary basis for processing web analytics data. If the site collects and respects user consent, the share of recorded data will be a function of the share of consenting users. Quantcast, the dominant GDPR consent management platform, reports that average website consent rates exceed 90%. As with privacy salience, we expect consenting users will self-select, so that we can we will detect differences in user quality metrics. [In the future, we may be able to obtain auxiliary data on whether and when sites implemented consent management platforms.]

Using the above logic, we look for evidence that speaks to the different mechanisms. However, firms differ in when and how they comply with the GDPR. As such, many explanations may coexist. Still, the above mechanisms suggest different policy ramifications: reducing the share of data recorded may be the more intended policy goal than reducing total web outcomes.

3 Data

On-site analytics are common tools used online to help firms better understand the types and activities of users on their site. While there are many vendors providing these solutions, the Adobe

Experience Cloud, of which Adobe Analytics is part, is one of the largest by market share and a leader in the field (Forrester 2017). Adobe Analytics both implements recording and tracking technologies for the firm as well as aggregates, cleans, and analyzes the collected data. Recording and tracking is executed primarily through java script. When a user arrives on site, code is triggered and a unique ID is assigned to the user-browser. All behavior during this sessions is then recorded at the interaction level and sent to Adobe's servers where it goes through various cleaning, assignment, and aggregation stages. Adobe allows each firm to assign privacy labels to individuals, and/or individual data fields, using its *Data Governance* interface. Privacy labels allow each firm to specify which data may be sensitive and need to be anonymized or deleted and which users do not wish to be recorded (Adobe 2018). The cleaned and aggregated data is then displayed to the firm in a *Report Suite* from which individual reports and views are constructed.

Our data consists of the aggregated user-web page interactions to a weekly panel of outcomes for a large number of global firms. We observe four key outcomes at the country-week level:

- **Page views** - Consists of a request for full page document by a visitor on site. This excludes partial requests, such as for a particular image or video.
- **Visits** - Consists of a sequence of consecutive page views without a 30 minute break.
- **Orders** - The number of a times a purchase event occurs.
- **Revenue** - Captured at the time of a purchase event and defined as the total currency amount for the sum of the order and each product.

For more technical information please see the Adobe Analytics reference documentation (Adobe (2018)). We aggregate these outcomes to the week European Union level. After aggregating, we are left with a weekly panel of the above four outcomes for 1508 analytics dashboards or *Report Suite Identifiers* (RSID's). It is important to note that this data is constructed in order to provide each firm with insight into its own online presence, operations, and customers - not for research purposes. In particular, RSID's may consist of arbitrary aggregations of traffic across multiple domains or sub-domains. For example, considering a retailer with both US and UK facing sites. The retailer could choose: (1) to combine both sites into one RSID, (2) create an RSID for each site, or (3) create multiple RSID's with duplicate information. We are able to filter out duplicated RSID's for our analysis, but not split RSID's up by domain - thus we take RSID as our natural unit of analysis.

The final data set is obtained at a weekly frequency, by RSID-country, for the 2nd through 38th weeks of 2017 and 2018. Because the implementation of the GDPR was on Friday, May 25th 2018, we define a week as a Friday-to-Friday period. Weekly aggregation is chosen to help mitigate high fluctuations in daily site traffic. The 2nd through 38th weeks avoid major holiday shopping and provide sufficient sample to effectively control for trends in the data.

Data is pulled by RSID; we start with a large list of RSID's and filter our sample down to the 1508 number above. Filtering is necessary as a large number of RSID's are used for testing & development, are non-current clients, and may duplicate data and lead to double counting in our analysis. All of the previous types of RSID's are filtered out. Furthermore, we filter out any RSID for which there are less than 100 average daily visits from the EU in the pre-treatment period. This is because with such low levels of European traffic, noise is an issue. RSID's for which the ratio of average daily visits (in the pre-treatment period) in 2017 versus 2018 is larger than 170 percent or smaller than 30 percent are dropped to avoid situations in which there is a change in reporting. Finally, we drop all RSID's that we can identify as corresponding to mobile dashboards, and for which there are serious reporting outages (more than 3 weeks of zero data in the pre-period). With the remaining RSID's we construct a panel at a weekly frequency for the dates above. This results in a final data set consisting of 1508 RSID's for which we have weekly data throughout the whole period above and meet the filtering criteria.

3.1 E-Commerce Sample Selection

There is an important distinction, with respect to reporting, between page views/visits and orders/revenue. While page views and visits are automatically collected and aggregated by the Adobe Analytics platform, orders and revenue (E-commerce outcomes) are only constructed when requested by the client. This is not ubiquitous, as many clients are not E-commerce firms. Similarly, many RSID's do not correspond to clients selling merchandise, thus such metrics are meaningless. Therefore, some filtering of the sample is necessary to not bias our E-commerce results towards

	All Firm Sample		E-Commerce Sample			
	Page Views	Visits	Page Views	Visits	Orders	Revenue
N	1,508	1,508	421	421	421	421
Mean	3,090,535.26	691,466.33	3,136,538.97	558,654.30	6,802.92	1,199,592.00
5th Percentile	3,285.94	1,329.44	9,350.71	2,543.43	14.52	1,863.78
Median	102,267.54	33,429.30	238,413.19	52,217.33	572.05	102,954.24
95th Percentile	11,638,668.22	2,450,527.72	11,073,652.29	2,866,793.10	35,073.81	5,409,354.66

Table 2: 2018 EU pre-treatment weekly summary statistics Our RSID's represent a substantial amount of E-commerce and are somewhat diverse in their size. The full sample of firms seem to be somewhat smaller than the E-commerce sample.

zero. In particular, we remove any RSID's for which we observe average daily revenue and orders of less than one[3] across the entire pre-treatment period. Fundamentally, due to the ad hoc nature of the revenue data collection, these metrics are subject to more noise than page views and visits. For this reason we additionally trim the distribution of revenue, only including the middle 1 - 99 percentiles. After filtering, this results in a sample of 421 RSID's.

3.2 Descriptive Results

Our final data set contains a significant portion of total E-commmerce revenue, roughly $15 billion per month in North America and $3 billion per month in Europe. This accounts for almost half of the estimated average spending in North America per month[4]. Additionally, it represents a large sample of web traffic; our sample of RSID's constitute approximately 63 billion page views per month. While these cumulative numbers are compelling, a significant strength of our data is the diversity of site types and sizes contained in it. Table 2 and figure 1 illustrate the variety of sites in our data. In table 2 we first calculate the average per week within each RSID and then calculate the summary statistic across RSID's. Comparing the median and 95th percentile, the largest firms in our sample are almost 100 times larger than the middle firms. This long tail motivates our preference for logged dependent variables in our analysis. Figure 1 plots the logged average weekly outcome of each RSID. We can see that these distributions look approximately normal, but still exhibit significant dispersion.

To get a sense of the types of firms our RSID's represent, as well as their size, we can take

[3] One is chosen rather than zero to exclude RSID's for which we see one or very few days of positive revenue.

[4] Annual E-commerce spending in the United States in 2017 was 453.5 billion, as estimated by the U.S. census. This is approximately 37.8 billion per month - not accounting for large increases in spending around Christmas and Thanksgiving which we excluding from our average monthly spending calculation above. Source: U.S. Census Quarterly E-Commerce Report

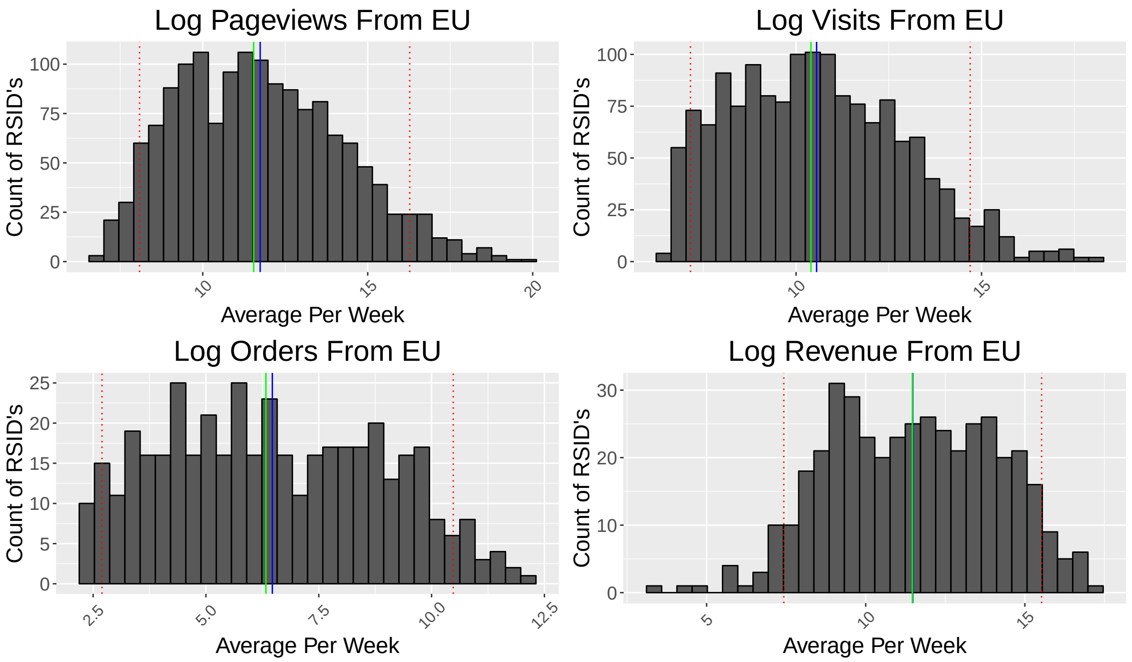

Figure 1: Heterogeneity in RSID logged outcomes The green lines mark the median of the distribution, blue lines the mean, and red dotted lines the 5th and 95th percentiles. Page views and visits are presented for the full sample.

advantage of site classifications from Amazon Web Information Services (AWIS). In particular, we can use the URL's associated with our RSID's to merge meta data from Adobe with AWIS. Of our 1508 RSID's we can successfully match 1293 to at least one URL in the AWIS top 1 million sites. Because each RSID may be associated with multiple URL's we present two metrics of site rank:

1. $\text{rank}(\text{RSID}_j) = \min_i \text{rank}(URL_i)$ s.t. $i \in \text{RSID}_j$
2. $\text{rank}(\text{RSID}_j) = \text{mean}_i \text{ rank}(URL_i)$ s.t. $i \in \text{RSID}_j$

Both of the above measures have short comings. Because RSID's often have both large and small sites contributing to them the first measure likely overstates the RSID's rank while the second understates it. From table 3 we can see that most of our RSID's represent relatively large sites in the top 10,000 sites globally and many correspond to multiple sites with diverse ranks. We also include the total number of URL's we cover within each bucket - our 1293 RSID's correspond to 2594 different URL's in the Alexa top million sites.

Of paramount interest for our empirical strategy are the intertemporal trends present in the data, pre and post GDPR. In figure 2 we present weekly averages, across RSID's, for the key metrics of interest across the sample period. The vertical line marks May 25th. To more clearly illustrate the trends in the data, data from the prior year is included as well. In general, there is

	min(rank)	avg(rank)	Site Count
< 1000	91	49	128
< 10000	419	268	595
< 100000	936	800	1592
< 1 million	1293	1293	2594

Table 3: RSID AWIS site rank Our RSID's are generally large and correspond to many sites of diverse sizes.

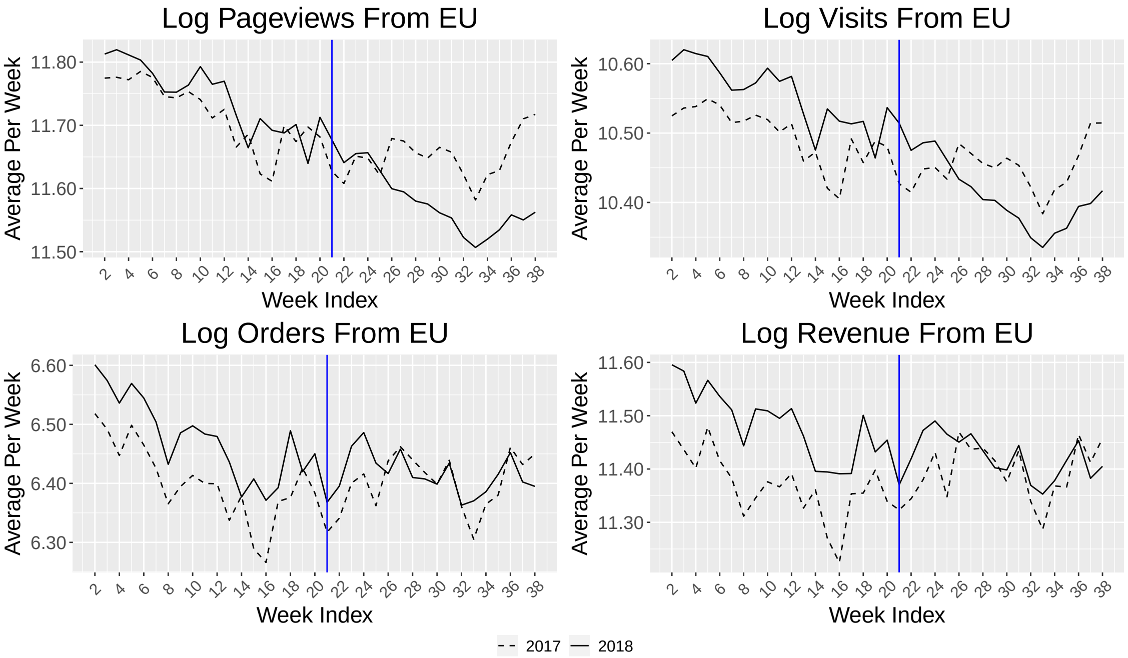

Figure 2: Average weekly outcomes 2018 trends of our key outcomes versus their 2017 counterparts. There is evidence that trends are similar in 2017 and 2018. One noticeable feature is the distinct level shift in 2018 outcomes in the post-GDPR period. Page views and visits are presented for the full sample.

evidence of a distinct level shift in the post period, roughly 4 weeks after the implementation of GDPR, relative to the trends in 2017. We can see that this trend is persistent for the remainder of the sample and that the behavior of the 2017 time-series is similar to that of the 2018 time-series.

One key feature of the above figures is the slightly delayed crossing of the 2018 and 2017 trends. In particular, this happens roughly four weeks after the GDPR is implemented. This lag could be driven by the cumulative effect of a decrease consumer tracking in targeting, and/or a heterogeneous timing of implementation of consent walls. While we do not have data on the timing of firm actions, section 2.1 discusses ways in which we can try to recover evidence of different firm behaviors from our data. Section 5.2 provides some evidence of robustness to potential anticipation and delay effects.

4 Differences-in-Differences Results: Adobe Analytics Data

4.1 Recorded Outcomes

The primary analysis will relies on a differences-in-differences specification. The ideal control group in this setting would be a set of RSID's for which a substantial portion of traffic is from the EU and who do not have to comply with GDPR post May 25th. As GDPR applies to all EU web traffic, this is infeasible. Instead, we propose using the same RSID 2017 traffic as our control group. This is a reasonable control group in that it effectively accounts for any seasonal or time effects orthogonal to yearly differences. For example, June is the beginning of the European summer for which our 2017 control group should effectively control for. One might also consider using contemporaneous outcome measures from North America as a control group. Unfortunately, this group is likely to suffer from substantial spill over effects from the GDPR as many organizations implemented their compliance solutions globally[5]. Figure 2 suggests that using 2017 same RSID metrics as our control group is viable. In section 5.4 we explore other potential control group specifications.

Our differences-in-differences regression takes the following form:

$$\log\left(y_{it}+1\right) = \alpha\mathbb{1}\{2018\} + \beta\left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}\right) + \theta_i + \eta_{iw} + \epsilon_{itw} \tag{1}$$

The primary coefficient of interest is β, which is an estimate of the average treatment effect under normal differences-in-differences identification assumptions (Angrist& Pischke 2008). To simplify notation in the following, t will refer to year, w to week, and i to RSID. Our recorded outcome is y_{itw}, θ_i are fixed effects at the RSID level, $\eta_i w$ are RSID x week fixed effects, $\mathbb{1}\{2018\}$ is a dummy variable for if the observation is from 2018 (analogous to the treatment-control dummy in a standard differences-in-differences specification) and $\mathbb{1}\{\text{Post GDPR}\}$ is a dummy variable for after the enforcement date of May 25th. Note that $\mathbb{1}\{\text{Post GDPR}\}$ dummy is collinear with the week fixed effects, and thus omitted. As discussed in section 3.2, we use logged outcome measures as the distributions of recorded outcome measures are highly skewed.

We can apply the above regression design to our metrics to examine how GDPR has impacted our four key outcomes. Results for our preferred specification are presented in table 4.

[5] https://qz.com/1284895/what-gdpr-compliance-means-for-american-businesses/

	Full Sample		*E-Commerce Sample*			
	Page Views	Visits	Page Views	Visits	Orders	Revenue
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{\text{Post GDPR}\}$	-0.097^{***}	-0.099^{***}	-0.042^{*}	-0.050^{**}	-0.056^{**}	-0.083^{***}
	(0.027)	(0.025)	(0.022)	(0.022)	(0.025)	(0.029)
$\mathbb{1}\{2018\}$	0.028^{*}	0.060^{***}	0.044^{**}	0.085^{***}	0.073^{***}	0.113^{***}
	(0.016)	(0.014)	(0.020)	(0.019)	(0.021)	(0.022)
RSID FE	Y	Y	Y	Y	Y	Y
RSID x Week FE	Y	Y	Y	Y	Y	Y
Observations	108,576	108,576	30,312	30,312	30,312	30,312
R^2	0.987	0.989	0.993	0.992	0.991	0.987
Adjusted R^2	0.974	0.978	0.986	0.985	0.983	0.974

Note: All DV's are logged. SE's clustered at RSID + Week level
*p<0.1; **p<0.05; ***p<0.01

Table 4: Differences-in-Differences Results: All Firm Sample & E-Commerce Sample.

We see large statistically significant decreases across all of our key metrics. For the full sample of firms, we see an average decrease of -0.097 and -0.099 for both visits and page views. These outcomes drop to -0.042 and -0.050 when only looking at our sample of firms with revenue outcomes. Our point estimate of 9 percent for visits suggests a weekly decrease of approximately 3,350 visits at the median of our all firm sample. Our point estimates for orders and revenue are a bit larger, at -0.056 and -0.083 respectively. These are large and economically consequential changes. The revenue results indicate approximately an \$8,000 drop in weekly recorded revenue for the median RSID in our sample. The precision of the estimates falls as we move towards revenue. As such caution should be used in over interpreting differences in point estimates.

4.2 Heterogeneity

Evidence presented in section 3 suggests there is significant heterogeneity across our sample of RSIDs. This will likely lead to heterogeneity in site outcomes as well. To examine this we can use a triple differences-in-differences strategy. We begin by classifying sites as either 'small' or 'large.' To do so, we split the distribution of average weekly visits, from the EU, in the pre-GDPR period (presented in figure 1) at the median - with sites above the median classified as 'large.' This is done separately for the for full firm sample and the E-commerce sample. We then estimate equation 2.

$$\begin{aligned}\log\left(y_{itw}+1\right) = \; & \alpha_1 \mathbb{1}\{2018\} + \alpha_2 \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Large}\}\right) \\ & + \alpha_3 \left(\mathbb{1}\{\text{Post GDPR}\} \text{ x } \mathbb{1}\{\text{Large}\}\right) \\ & + \beta_1 \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}\right) \\ & + \beta_2 \left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\} \text{ x } \mathbb{1}\{\text{Large}\}\right) \\ & + \theta_i + \eta_{iw} + \epsilon_{itw} \qquad (2)\end{aligned}$$

From equation **??** we are interested in the differences between β_1, the treatment effect for small firms, and β_2, the treatment effect for larger firms. All

4.3 User Behavior

4.3.1 Visit Quality Metrics

Though we find large changes in recorded outcomes in table 4, we see no such changes in visit quality metrics. To examine this we construct two metrics commonly used by site managers and practitioners, time per visit and page views per visit, and use them as our outcome variables in equation 1. The trends and results of these regressions are presented in figure 3 and table 5, respectively.

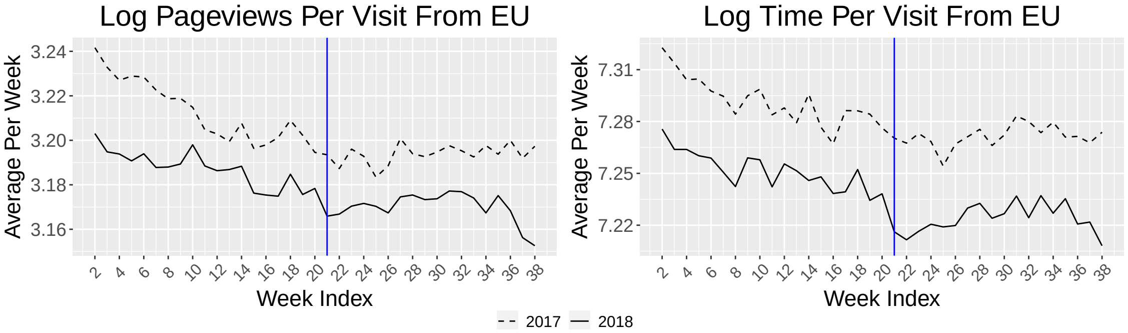

Figure 3: Average weekly outcomes: quality metrics Blue vertical lines indicate the date of treatment. There is little evidence of differences in trend across years, or in the post-GDPR period.

Appealing to figure 3, we see that both metrics have higher levels across the pre and post period in 2017. In both cases, there is little evidence of change in the post GDPR period. Table 5 demonstrates that we see no evidence of a change in how much users are using sites, on average. Importantly, these estimates are precise - our standard errors suggest that were effects large enough to be economically meaninful, we would be able to measure them. These null effects suggest that the types of users who visit in the post period are similar to those that were visiting in the pre-period.

	Dependent variable:	
	Page Views Per Visit	Time Spent Per Visit
	(1)	(2)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{\text{Post GDPR}\}$	0.004	-0.003
	(0.006)	(0.010)
$\mathbb{1}\{2018\}$	-0.020^{***}	-0.045^{***}
	(0.006)	(0.010)
RSID FE	Y	Y
RSID x Week FE	Y	Y
Observations	108,576	108,576
R^2	0.950	0.945
Adjusted R^2	0.899	0.891

Note: All DV's are logged. SE's clustered at RSID + Week level
*p<0.1; **p<0.05; ***p<0.01

Table 5: User Behavior Regression

To the extent that we think different mechanisms may impact the distribution of consumer types who visit a site, the above statistics are informative. In particular, because we see little change in how much time users spend on a site per visit, we take this as evidence that privacy salience is playing a limited role. On the contrary, if privacy salience was leading users to change their browsing behavior, we would likely see a positive effect. Users who value the site less would substitute away from using the site, leading average quality statistics to increase as the remaining population values the site, on average, more. On the other hand, if the utility from site usage and privacy concerns are positively correlated, this argument fails. It is unclear under what circumstance privacy preferences and site utility may be positively correlated, but we cannot rule this out. In this case, we might expect users to utilize the opt-out features made available by the GDPR - leading to high-value users dropping out of our data and our quality metrics decreasing in the post-GDPR period. We also do not see this, though we cannot rule out that the two effects above cancel each other out. Therefore, we conclude that privacy salience and opt-outs are either playing a limited role or uncorrelated with privacy preferences.

5 Robustness

5.1 Placebo Tests

While the point estimates we present in table 4 are large and statistically significant, they do not rule out the potential of false positives. In order to address this we can run placebo tests. We implement placebo tests by first choosing a counterfactual treatment week from the pre-GDPR period of our data. Then equation 3 is estimated using data from before April 25th (pre-GDPR). We use exclude one-month before the implementation of the GDPR in order to omit any anticipation behavior. This procedure is repeated for placebo treatment dates ranging from 17 to 7 week prior to May 25th for a total of 11 placebo tests. These placebo dates are chosen in order to provide adequate pre-trend and post-trends in the data (at least 3 data points before and after the placebo treatment).

$$\log\left(y_{itw}+1\right)=\alpha\mathbb{1}\{2018\}+\beta_p\left(\mathbb{1}\{2018\}\text{ x }\mathbb{1}\{\text{Post Placebo}\}\right)+\theta_i+\eta_{iw}+\epsilon_{itw} \tag{3}$$

The primary coefficient of interest is β_p. Fixed effects are included just as in equation 1 and all standard errors are clustered at the RSID + week level. Significant point estimates are indicative of false positives, and may undermine the credibility of our point estimates in table 4. Figure ?? plots the estimates and confidence intervals for each of the placebo tests.

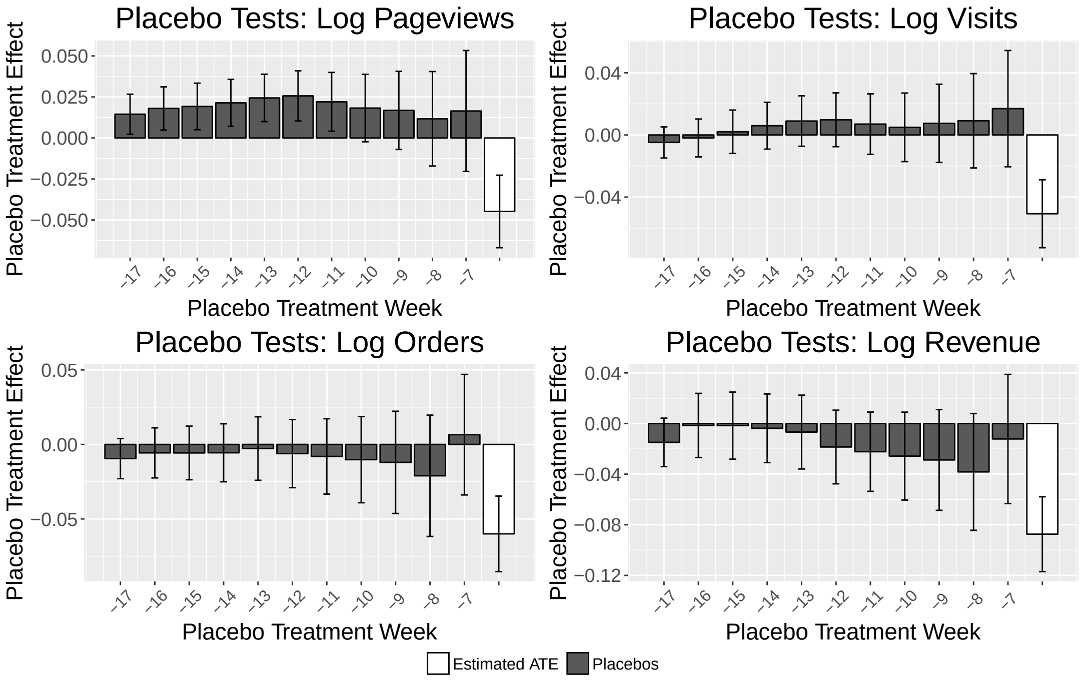

Figure 4: Placebo tests Each bar corresponds to the point estimate of the placebo average treatment effect. Bars are standard errors - clustered at the RSID + week level. The white bar indicates the ATE estimated in table 4

The placebo results generally demonstrate a robustness of our identification strategy and point estimates in table 4. We can see that no estimates match the magnitude of our main results, nor are any of the placebo tests nearly as statistically significant. Confidence intervals tend to get larger as we move to the right in each of these figures - this is a mechanical reflection of the decreasing number of post-placebo treatment periods. While page views have a substantial number of significant treatment effect estimates, they are positive and generally small point estimates. These significant effects, as well as the trend in placebo estimates for orders and revenue, may be suggestive of a poor control group, which leads us to considering other control groups in sections 5.4 and 5.3.

5.2 Timing of Implementation

A critical assumption to identification of the true average treatment effect in 4.1 is that there is not pre May 25th 2018 implementation of GDPR compliance, or no treatment anticipation. While we do not directly observe compliance behaviors, anecdotal evidence resoundingly suggests that firms were unprepared for the GDPR [6]. In order to address treatment anticipation we can remove a one month window around our treatment date and re-run the regression in equation 1. This exercise will help remove any transitory effects around the implementation of the GDPR. Results of this exercise are presented in table 6.

	Full Sample		*E-Commerce Sample*			
	Page Views	Visits	Page Views	Visits	Orders	Revenue
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{\text{Post GDPR}\}$	-0.095^{***}	-0.096^{***}	-0.044^{**}	-0.049^{**}	-0.054^{**}	-0.079^{***}
	(0.028)	(0.026)	(0.022)	(0.022)	(0.026)	(0.030)
$\mathbb{1}\{2018\}$	0.026	0.056^{***}	0.045^{**}	0.083^{***}	0.071^{***}	0.109^{***}
	(0.017)	(0.015)	(0.021)	(0.020)	(0.022)	(0.024)
RSID FE	Y	Y	Y	Y	Y	Y
RSID x Week FE	Y	Y	Y	Y	Y	Y
Observations	99,528	99,528	27,786	27,786	27,786	27,786
R^2	0.987	0.989	0.993	0.992	0.991	0.987
Adjusted R^2	0.973	0.977	0.985	0.984	0.982	0.973

Note: All DV's are logged. SE's clustered at RSID + Week level
*p<0.1; **p<0.05; ***p<0.01

Table 6: Window Regressions

Comparing tables 6 and 4, we can see that our point estimates are unchanged. Thus, it seems that at least short term anticipation effects are not playing a large role in our identification strategy.

5.3 Triple Differences-in-Differences

If global macroeconomic trends differ substantially in 2018 and 2017 our regression in equation 1 will not estimate the true average treatment effect. One way to address this is through a triple differences-in-differences regression design using our North America data to control for global changes in trends. This strategy is also potentially flawed; to the extent that there are spill-overs from the GDPR on North American outcomes, and these spill-overs negatively influence outcomes, regression 1 will underestimate the average treatment effect.

[6] https://www.theguardian.com/technology/askjack/2018/jul/05/what-should-i-do-about-all-the-gdpr-pop-ups-on-websites

We estimate the following equation:

$$\begin{aligned}\log\left(y_{itwn}+1\right) = \ & \alpha_1 \mathbb{1}\{2018\} + \alpha_2\left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{EU\}\right) \\ & + \alpha_3\left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\}\right) \\ & + \beta\left(\mathbb{1}\{2018\} \text{ x } \mathbb{1}\{\text{Post GDPR}\} \text{ x } \mathbb{1}\{EU\}\right) \\ & + \theta_i + \eta_w + \xi_{inw} + \nu_{ni} + \epsilon_{itwn} \end{aligned} \tag{4}$$

The regression is run at the RSID (i), year (t), week (w), region (n) level. The primary coefficient of interest is β. We include full interactions of indicators for geography, year, and post-GDPR. Note that indicators for post-GDPR, EU, and post-GDPR x EU are absorbed by fixed effects; we include RSID, week, week x region, and region x RSID fixed effects. Estimation results from equation 4 are presented in table 7.

As expected, we can see that the majority of our point estimate decrease but remain substantially negative and significant. The one exception to this our E-commerce sample estimate for page views, which is very small and insignificant.

5.4 Synthetic Controls

While previous year same RSID web traffic may help control for seasonality there are potentially other concerns in using this approach. In particular, changes in reporting across years, changes in web traffic due to advertising or product launches, and other cross-year within-firm differences could lead to a violation of the parallel trends assumption necessary for identification in a differences-in-differences model. In order to provide evidence of robustness of our results to such concerns we appeal to the method of synthetic controls (Abadie (2010) and Doudchenko & Imbens (2017)).

Synthetic controls constructs a control group by taking a weighted average of control units to best predict the counter factual for the treated unit. This control group is known as the 'synthetic control' and will satisfy parallel trends *by construction.* With a control and treatment group in hand, we can recover the average treatment effect through an approach such as differences-in-differences. Intuitively, the idea is that if we can construct a control group that behaves similarly enough to the treatment group in the pre-period, then this control group behaves similarly to how

	Full Sample		*E-Commerce Sample*			
	Page Views	Visits	Page Views	Visits	Orders	Revenue
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{\text{Post GDPR}\}$ x $\mathbb{1}\{EU\}$	-0.028^{***}	-0.022^{***}	-0.001	-0.010^{***}	-0.038^{***}	-0.054^{***}
	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{\text{Post GDPR}\}$	-0.070^{***}	-0.077^{***}	-0.041^{***}	-0.040^{***}	-0.017^{*}	-0.029^{***}
	(0.013)	(0.013)	(0.007)	(0.007)	(0.009)	(0.011)
$\mathbb{1}\{2018\}$	0.051^{***}	0.122^{***}	0.080^{***}	0.180^{***}	0.023^{**}	0.008
	(0.008)	(0.006)	(0.008)	(0.007)	(0.009)	(0.010)
$\mathbb{1}\{2018\}$ x $\mathbb{1}\{EU\}$	-0.023^{***}	-0.062^{***}	-0.036^{***}	-0.095^{***}	0.049^{***}	0.105^{***}
	(0.003)	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
RSID FE	Y	Y	Y	Y	Y	Y
RSID x Week FE	Y	Y	Y	Y	Y	Y
Region x RSID FE	Y	Y	Y	Y	Y	Y
RSID x Region x Week FE	Y	Y	Y	Y	Y	Y
Observations	217,127	217,127	60,603	60,603	60,603	60,603
R^2	0.988	0.988	0.993	0.992	0.993	0.977
Adjusted R^2	0.975	0.977	0.987	0.984	0.987	0.954

Note: All DV's are logged. SE's clustered at RSID + Week + Region level

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

Table 7: Triple Differences-in-Difference Regression

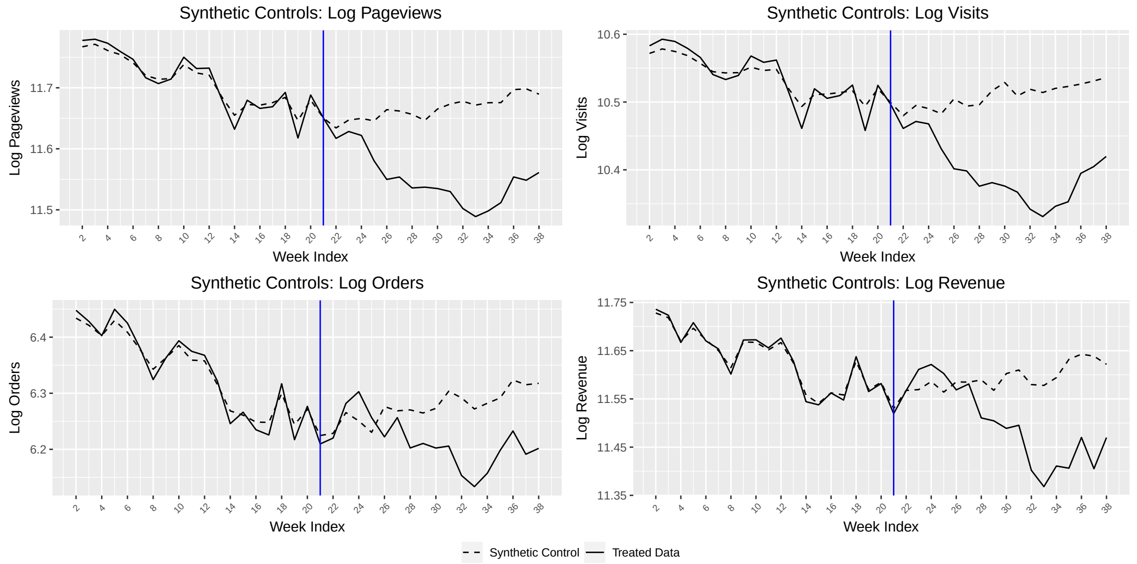

Figure 5: Fitted Synthetic Controls. We can see a good (as expected) fit in the pre-GDPR period and reasonable trends in the post period that seem overall quite similar to those noted in figure 2. Page views and visits are presented for the full sample.

the treatment group would have behaved, after the intervention date, had it not received treatment. The problem then becomes one of constructing the synthetic control. In our implementation, we use 2017-RSID data as our control units and the mean of 2018-RSID data as our treated unit. Because we have many more control units than pre-treatment periods, we follow Doudchenko & Imbens (2017) and use an elastic net to construct our synthetic control. Details on the cross-validation and model fitting procedure are presented in appendix A.

We begin by plotting the fitted trends in figure 5. We can see that our elastic net does a good job of fitting our treated group in the pre-GDPR period and trends in the post period are largely similar to the average 2017 trends in figure 2. Note that we see large drops discrepancies between the synthetic control group and the treated group in the post-GDPR period - indicating that the GDPR may have had an effect. Similarly, we can see that the delayed crossing noted in figure 2 is apparent for the orders and revenue outcomes here. This may not be surprising as the synthetic control group is constructed of 2017 RSID data - but does suggest that trends in 2017 in the post-GDPR period are substantially different than in 2018. Our point estimates for each outcome are presented in table 8. These estimates are generally in line with the magnitudes presented in table 4 are demonstrate some robustness to our same trends assumption.

While there is no universal theory of inference for synthetic controls, particularly when

	DID Point Estimate	Placebo Dist. Quantile
Page Views	-0.116	24
Visits	-0.100	20
Orders	-0.067	34
Revenue	-.033	30

Table 8: Synthetic Controls Results: This table presents the level differences-differences estimates and quantile in the placebo distribution of each outcome

elastic nets are used to fit the control group, we can appeal to Abadie (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 $n = 25$ units from C
- Construct our psuedo-treated unit as $C_t^{psuedo} = \frac{1}{n}\sum_{i \in samp} C_{it}$
- Fit an elastic net to C_t^{psuedo} as described in appendix A
- Calculate the adjusted mean squared prediction error (Abadie (2010)):

$$\frac{T_0}{T - T_0} \frac{\sum_{t=T_0}^{T} \left(Y_t^{psuedo}(1) - Y_t^{C(psuedo)}(0)\right)^2}{\sum_{t=0}^{T_0} \left(Y_t^{psuedo}(1) - Y_t^{C(psuedo)}(0)\right)^2} \tag{5}$$

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

- Repeat 250 times for each outcome

The above procedure gives us a sense of how large the adjusted mean square error calculated using the actual treated units (the synthetic controls reflected in figure 5) is compared to the same procedure on non-treated units. Our takeaways from this procedure should be that if the adjusted mean squared error in the treated case is larger than in the psuedo-treated case, there is some evidence that our treatment has had an effect. Importantly, this procedure will disadvantage our estimates as using an $n = 25$ to construct our psuedo-treated unit will lead to a noisier treated group than in the actual case, in which we use the full sample of firms. The results of the above procedure are presented in figure 6.

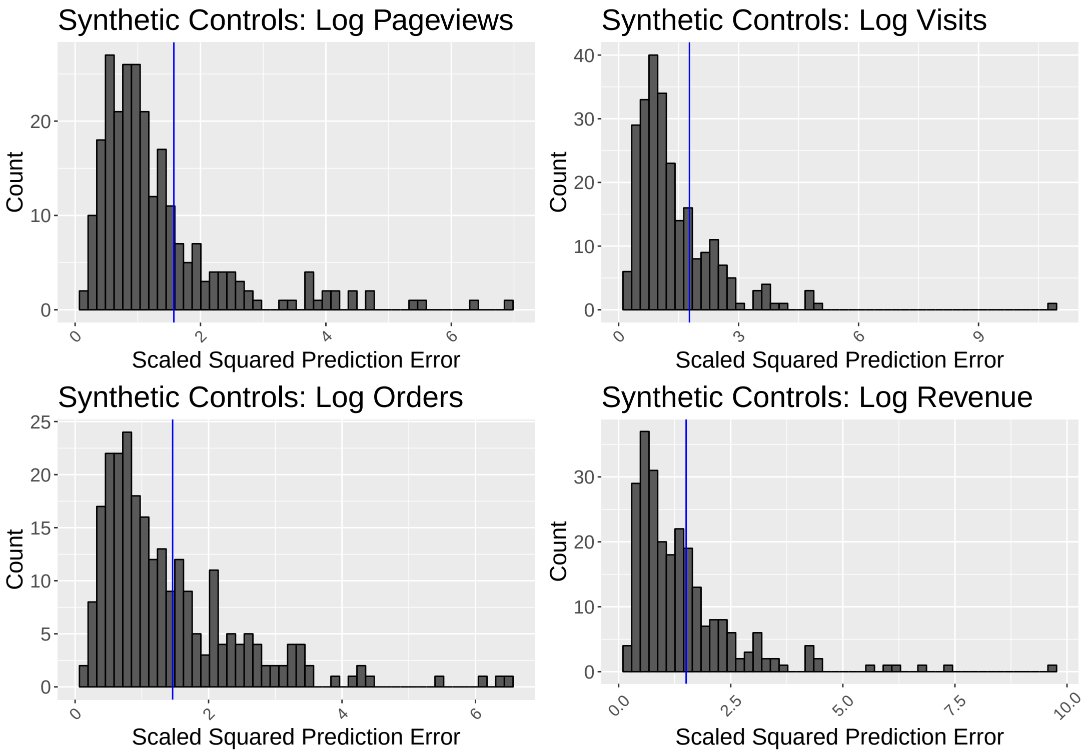

Figure 6: Synthetic Control Placebo Tests The blue line is the scaled squared prediction error of the true treated cased. Page views and visits are presented for the full sample.

Figure 6 suggests that our results are somewhat robust - though there is clearly some noise in this data. The quantile in the placebo distributions or our true synthetic controls estimate is presented in table 8. Generally the true estimates are in the upper end of the placebo distributions but not conclusively outliers. This may indicate that interpreting figure 5 as indicating a *large* GDPR impact on revenue or orders is overstating the true effect. Nonetheless there is evidence that these effects are substantial and larger than random.

6 Conclusion

This paper has attempted to begin to answer a complicated and nuanced question: how has the European Union's General Data Protection Regulation impacted online outcomes? To answer this question we have used an expansive and comprehensive new data set.

Using data from Adobe Analytics, we are able to quantify the impact of GDPR on important economic outcomes for a diverse set of firms. We find large mean effects: page views per week drop by approximately 4% and revenue per week falls by 8%. These are economically large numbers, with a 8% revenue per week drop corresponding to a \$8,000 drop in weekly revenue for the median RSID in our sample. We provide some evidence that these results are not driven by changes in user behavior directly. From a regulators perspective the above results clearly illustrate the difficulty and high costs of privacy regulation. The Adobe Analytics data illustrate just a portion of the total cost of complying with GDPR - omitting large operational and infrastructure costs. More work needs to be done to quantify the benefit to users of these privacy laws in order to better understand the tradeoffs. An indicator of this are the opt-in numbers cited in the introduction, which suggest that GDPR may not actually be delivering that much value to the majority of users. Regardless, it is likely true that GDPR has impacted different sites in very different ways, and regulators may want to consider this when working on future legislation. Firms have reacted in diverse ways and have implemented compliance in various fashions, often driven by their business needs. Legislators may want to consider why and how firms are using user information more explicitly in legislation to better address these asymmetries.

While this paper has some interesting and compelling findings more work is left to be done. In particular our next steps include using last touch attribution data to examine if users have changed how they arrive on site and bringing in auxiliary data to more directly examine the data minimization and consent channels of the mechanism.

7 Appendix

A Cross-validation Routine for Synthetic Controls

Estimating the weights necessary to construct the control group is difficult and not well understood (Doudchenko and Imbens (2017)). 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 RSID log outcome data as control units and the average across RSID's 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:

$$Y_t = \frac{1}{N}\sum_{i}^{N} \log\left(y_{it}^{2018} + 1\right) \tag{6}$$

$$C = \{C_{it} = \log\left(y_{it}^{2017} + 1\right) \forall\, i\} \tag{7}$$

The problem of estimating weights then becomes one of choosing a weighted combination of C_{it} to best match Y_t. The literature has used a variety of different methods to accomplish this task (Doudchenko and Imbens (2017)). In our setting, we have twenty-one pre-treatment time periods and 1508 control units. That is, we have a situation in which $N >> T_0$. Because of this, we follow Doudchenko and Imbens (2017) 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:

$$Q(\mu, \omega | Y_t, C_{it}; \alpha, \lambda \text{ for } t < T_0) = ||Y_t - \mu - \omega C_{it}||_2^2 + \lambda \cdot \left(\frac{1-\alpha}{2}||\omega||_2^2 + \alpha||\omega||_1\right) \tag{8}$$

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 (2017).

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 $i \in C^b$. We use $\tilde{C} = C \setminus C^b$ 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:

$$\{\hat{\mu}^b, \hat{\omega}^b\} = \text{argmin}_{\mu,\omega} \sum_{t=1}^{T_0} \left(Y_t^{C^b} - \mu - \omega\tilde{C}_{it}\right)^2 \; + \; \lambda' \cdot \left(\frac{1-\alpha'}{2}||\omega||_2^2 + \alpha'||\omega||_1\right) \tag{9}$$

Given the weights estimated above, 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.

$$Y_t^{C^b}(0) = \hat{\mu}^b + \hat{\omega}^b \tilde{C}_{it} \tag{10}$$

$$CV_B(\alpha', \lambda') = \frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(Y_t^{C^b}(1) - Y_t^{C^b}(0) \right)^2 \tag{11}$$

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

$$CV(\alpha', \lambda') = \frac{1}{B} \sum_b CV_b(\alpha', \lambda') \tag{12}$$

Finally, tuning parameters are chosen such that $\{\alpha, \lambda\} = argmin_{\alpha', \lambda'} CV(\alpha', \lambda')$. Using these tuning parameters, the model in equation 8 is fitted with the control units and treatment groups constructed in equation 6. From here, a differences-in-differences regression is run to recover the average treatment effect under the assumption that $Y_i^C(0) = Y_i^T(0)$.

For our purposes, 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.

References

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

[2] Abadie, A., Diamond, A., and Hainmueller, J. Comparative politics and the synthetic control method. *American Journal of Political Science 59*, 2 (2015), 495–510.

[3] Acquisti, A., John, L. K., and Loewenstein, G. What is privacy worth? *The Journal of Legal Studies 42*, 2 (2013), 249–274.

[4] Acquisti, A., Taylor, C., and Wagman, L. The economics of privacy. *Journal of Economic Literature 54*, 2 (June 2016), 442–92.

[5] Adobe. Adobe experience cloud analytics help and reference. Tech. rep., Adobe, 2018.

[6] Athey, S., Catalini, C., and Tucker, C. The digital privacy paradox: Small money, small costs, small talk. NBER Working Paper No. 23488, June 2017.

[7] Bellman, S., Johnson, E. J., Kobrin, S. J., and Lohse, G. L. International differences in information privacy concerns: A global survey of consumers. *The Information Society 20*, 5 (2004), 313–324.

[8] Bennett, C. J., and Raab, C. D. *The governance of privacy: Policy instruments in global perspective.* 2006.

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

[10] Cline, J. Pulse survey: Gdpr budgets top 10 million for 40 percent of surveyed companies.

[11] Consumers Union. Poll: Consumers concerned about internet privacy. http://consumersunion.org/news/poll-consumers-concerned-about-internet-privacy/, September 2008. (Accessed on 07/19/2017).

[12] Creamer, M. Despite digital privacy uproar, consumers are not opting out. *Ad Age* (May 2011).

[13] de Matos, M. G., and Adjerid, I. Consumer behavior and firm targeting after gdpr: The case of a telecom provider in europe. Working paper, 2019.

[14] Doudchenko, N., and Imbens, G. Balancing, regression, differences-in-differences and synthetic control methods: A synthesis. NBER Working Paper Series No. 22791, October 2016.

[15] Fuller, C. S. Privacy law as price control. *European Journal of Law and Economics 45*, 2 (Apr 2018), 225–250.

[16] Gellman, R. Privacy, consumers, and costs-how the lack of privacy costs consumers and why business studies of privacy costs are biased and incomplete. In *Digital Media Forum, Ford Foundation* (2002).

[17] Godinho de Matos, M., and Adjerid, I. Consumer behavrio and firm targeting after gdpr: The case of a telecom provider in europe. Working Paper, May 2018.

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

[19] Goldfarb, A., and Tucker, C. Privacy and innovation. Working Paper 17124, National Bureau of Economic Research, June 2011.

[20] Goldfarb, A., and Tucker, C. Shifts in privacy concerns. *American Economic Review 102*, 3 (May 2012), 349–53.

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

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

[23] Gross, R., Acquisti, A., and John Heinz III, H. Information revelation and privacy in online social networks (the facebook case). In *WPES'05: Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society* (01 2005), pp. 71–80.

[24] Jia, J., Jin, G., and Wagman, L. The short-run effects of gdpr on technology venture investment. Working Paper, November 2018.

[25] Johnson, E. J., Bellman, S., and Lohse, G. L. Defaults, framing and privacy: Why opting in $\neq$ opting out. *Marketing Letters 13*, 1 (2002), 5–15.

[26] Johnson, G., Lewis, R., and Nubbemeyer, E. The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Working Paper, June 2017.

[27] Johnson, G., Shriver, S., and Du, S. Consumer privacy choice in online advertising: Who opts out and at what cost to industry? Working Paper, 2018.

[28] Johnson, G. A., Lewis, R. A., and Reiley, D. When less is more: Data and power in advertising experiments. *Marketing Science 36*, 1 (2017), 43–53.

[29] Johnson, J. P. Targeted advertising and advertising avoidance. *The RAND Journal of Economics 44*, 1 (1 2013), 128–144.

[30] Kim, T., Barasz, K., and John, L. K. Why am i seeing this ad? the effect of ad transparency on ad effectiveness. *Journal of Consumer Research* (2018), ucy039.

[31] Libert, T. An automated approach to auditing disclosure of third-party data collection in website privacy policies. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web* (2018), International World Wide Web Conferences Steering Committee, pp. 207–216.

[32] Marotta, V., Abhishek, V., and Acquisti, A. Online tracking and publishers' revenues: An empirical analysis. Working paper, 2019.

[33] McCormick, J., Leganza, G., and Miller, E. The forrester wave: Web analytics, q4 2017.

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

[35] Montes, R., Sand-Zantman, W., and Valletti, T. The value of personal information in online markets with endogenous privacy. *Management Science 65*, 3 (2019), 1342–1362.

[36] Pew Research Center. Public perceptions of privacy and security in the post-snowden era. Tech. rep., Pew Research Center, November 2014.

[37] Posner, R. A. The economics of privacy. *The American Economic Review 71*, 2 (1981), 405–409.

[38] Raine, L., Kiesler, S., Kang, R., and Madden, M. Anonymity, privacy, and security online. Tech. rep., Pew Research Center, September 2013.

[39] Shiller, B., Waldfogel, J., and Ryan, J. The effect of ad blocking on website traffic and quality. *The RAND Journal of Economics 49*, 1 (2017), 43–63.

[40] TRUSTe. 2011 consumer research results: Privacy and online behavioral advertising. Tech. rep., TRUSTe, 2011.

[41] Tsai, J. Y., Egelman, S., Cranor, L., and Acquisti, A. The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research 22*, 2 (2011), 254–268.

[42] Tucker, C. E. Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research 50*, 5 (2013), 546–562.

[43] Varian, H. R. Economic aspects of personal privacy. In *Internet Policy and Economics*. Springer, 2009, pp. 101–109.