Regulatory spillovers and data Governance: Evidence from the GDPR- a replication practice

Zih-Yun Yan & Jinyi Yang

Introduction

On 2018 May 25, GDPR (General Data Protection Regulation), the most rigid website privacy law that lets people request their data and restricts businesses from obtaining users' information enacted in the EU. As a leading global privacy regime, researchers would like to examine how this law empirically changed the behavior of the website, web technology providers, and the audience of the websites. To answer this question, the researchers followed more than 110,000 websites from May 2017 to November 2018, which covers data from 12 months before and six months after the GDPR came into force to see whether this law made any short-term changes and whether these changes would continue in the long run.

The primary principle of GDPR is *data minimization*, meaning personal data collection needs to be limited to what is necessary for legitimate processing purposes. Thus, the reduction of website firms requesting web technologies to collect and share personal data from third-party vendors can be an ideal setting to observe the website firms' responses to privacy regulation. For example, cookies, a web technology used to store on the user's local machine to recognize the user's preferences and other information even if the IP address is changed for that user, is an informative indicator for websites' reaction to the privacy law. Another advantage of using this web technology request from third-party vendors is that this information is publicly available to access.

Another empirical issue of this study is to understand the territorial application of GDPR. It can be applied to websites and website technology vendors in different cases. In addition to the standard case (case 1), both websites and users are in the EU. It is also designed to be in effect when the website is inside the EU but users accessing it are outside the EU (case 2). The third case it can be applied is when users are located inside the EU and access website outside the EU which still processes their data (case 3). The only situation where GEPR does not *legally* apply is when both websites and users are not in the EU. However, they find some empirical evidence that websites might still obey GDPR rules in this situation.

There are three significant findings shown in the data from this paper. First, they found that the website reduced the number of requests to third-party domains and cookies in all 4 cases. This finding suggests that the EU has shaped the privacy standard outside the legal boundaries of the EU and exported the regulatory framework to other countries. Second, there was an immediate decrease in third-party domain requests after GDPR. However, an upward trend of the requests across time can still be observed. However, a specific request, cookies, used to interact with consumers' privacy showed a sustained decrease after GDPR. The change in the

interactions between websites and third-party vendors is especially pronounced in vendors disclosing that they collected personal data. The third key finding from this paper is that even though the overall web technologies market of third-party vendors decreased due to the website sending fewer requests after the introduction of the GDPR, the dominant third-party vendors, such as Google, have increased in their shares among all web technologies market. This finding indicates that privacy regulation produces an unintended consequence for market structure because of the increase in compliance risk from the websites. Apart from the main results, the researchers also examined how the website's popularity affects the impact of the GDPR and whether the Facebook scandal event can be another confounding driver of their findings. In this paper, we replicated all conclusions and brought out two extensions to discuss the time trend effect of their results.

Data

The data for an investigation comes from both public and proprietary data sources. The primary dataset is from a documentation project named *HTTP Archive.* This dataset contains historical information about websites' HTTP requests to third parties. This paper used data that included 110,706 hosts over 33-time points (21 before GDPR and 12 after GDPR) from May 2017 to November 2018. The more extended pre-GDPR period is used as a robustness check. They collect information about the total number of requests to third-party domains, the identity of these domains, and the number of domains that respond to a request by sending cookies. In total, they have 3,653,2988 observations.

They used *Crunchbase* and *WHOIS* to obtain information about the website's headquarters location and country information. They observe country information from *Crunchbase* for 34,950 (31.6%) websites and 94,919 (85.77%) websites from *WHOIS* in their sample. To access Website audience information, they used country-specific top-level domains, EU-specific languages, and country-specific traffic to infer the website audience. The first two pieces of information were obtained from *HTTParchive*. The country-specific traffic information (based on page views and ranked in different countries) is from *Alexa.* After applying the territorial scope, they defined it based on their information. They have 19.4% of sample f websites belonging to EU firm-EU audience (case 1), 11.7% EU firm, non-EU audience (case 2). 6.8% to non-EU firms, EU audience (case 3), and 61.9% to non-EU firms and non-EU audience (case 4).

The technology vendor information was obtained from W*hotracks.me*. *Whotracks.me* also has the service categories information. This paper mainly focuses on five categories (audio/video player, advertising, analytics, CDN, and others) as these are the most important categories in terms of market share for Google in their sample.

Finally, they obtained information about whether a web technology provider discloses in its privacy policy whether it collects and shares data with third parties from Evidon's industry directory.

Models

In this section, we showed how researchers concluded their findings.

$$
\begin{aligned}
&Log(D_{it}+1) = \beta_1(Trend_t xEU - Audience) \\
&+\beta_2(Trend_t xnonEU - Audience) \\
&+\gamma_1(Trend_t xPost_t xEU - Audience_i)+ \\
&+\gamma_2(Trend_t xPost_t xnonEU - Audience_i)+ \\
&+\delta_1(Post_t xEU - Audience_i) + \delta_2 \\
&+(Post_t xnonEU - Audience_i) + \mu_i + \epsilon_{it}
\end{aligned} \quad (1)
$$

The basic model specification of their main finding is shown in equation (1). D_{it} is the number of third-party domains that website i requested at time t. In the similar specification, they also look at the subset of third-party domains that respond with at least one cookie. $Post_t$ indicates the period after the GDPR came into force on May 25, 2018 (1 after May 25, 2018, and 0 before May 25, 2018). The model also included group- and time period–specific linear time trends and website fixed effects μ_i. The error term ε_{it} has the standard assumptions, and they reported estimates clustered at the website level. In total, they created eight different regression models (four cases x third party requests and four cases x cookies requests) based on the equation (1).

$$
\begin{aligned}
&Log(D_{it}+1) = \beta_1(Trend_t xEU - Audience) \\
&+\beta_2(Trend_t xnonEU - Audience) \\
&+\gamma_1(Trend_t xPost_t xEU - Audience_i)+ \\
&+\gamma_2(Trend_t xPost_t xnonEU - Audience_i)+ \\
&+\delta_1(Post_t xEU - Audience_i) + \delta_2 \\
&+(Post_t xnonEU - Audience_i) \\
&+\delta_3(Post_t xEUAudiencexTop1k) \\
&+\delta_4(Post_t xnonEU - AudiencexTok1k) + \mu_i + \epsilon_{it}
\end{aligned} \quad (2)
$$

In addition to the equation (1), they further examined the effect of website popularity. In equation (2) , they included two other variables $Post_t$ *x EU-Audience x Top1k* and $Post_t$ *x nonEU-Audience x Tok1k* to see. The distinct website popularity to its global Alexa rank, which is based on overall traffic to understand the heterogeneity influence of GDPR on the websites. Five hundred ninety-nine websites are within the top 1000 most popular in their sample. In total, they created four regression models for each case.

$$
\begin{aligned}
&Log(D_{it}+1) = \\
&Post_t + trend_t + Post_t xCollectPersonal + trend_t xCollectPersonal \\
&+Post_t xSharePersonal + trend_t xSharePersonal \\
&+Post_t xCollectPersonalxSharePersonal \\
&+trend_t xCollectPersonalxSharePersonal
\end{aligned} \quad (3)
$$

In equation (3), they changed perspectives to focus on web technology vendors and to examine whether the changes they have observed and quantified from previous models are especially pronounced among vendors disclosing that they collect and share personal data.

$$HHI = After + trend + trend_{after} xAfter \quad (4)$$

In equation (4), they investigated changes in the market structure. They calculated the market structure of Google and non-google third-party vendors from the Herfindahl-Hirschman Index (HHI). After indicating the period after May 25, 2018. They also included separate linear time trends before/after the GDPR. They performed this analysis because relevant literature suggested that privacy protection can affect firms of different sizes differently, affecting the market structure.

$$
\begin{aligned}
&Log(D_{it}+1) = \\
&trend_t xGoogle + trend_t xnonGoogle \\
&+trendafterxafterxGoogle + trendafterxafterxnonGoogle \\
&+afterxGoogle + afterxnonGoogle
\end{aligned} \quad (5)
$$

In equation (5), after they have established that Google is the dominant firm in the market and that changes in aggregate market concentration largely on Google from previous regression analysis, they further compared the changes in the market share of Google to changes in the market shares of all other firms. The dependent variable is the number of websites served by all web technology firm services. They also used this Equation to examine market shares which resulted in 2 regression models. Finally, using the exact specification in Equation (5), they examined the number of websites that send requests to services owned by web technology firm level by submarket categories (Video, Advertising, Analytics, CDN/API and others) with five different regression models.

The figures and regression results were produced using Stata in the original paper. In our replication practice, we used Python to replicate their findings. Across five equation specifications, there were, in total 20 regression models.

Replication Experiments

In a nutshell, we completely replicated all regression models. As the researcher provided Stata code and their organized dataset for every regression model then ran. The code and regression summary tables are in the appendix. Here we showed the regression coefficient in interest and standard errors.

Replication table 1

Table 4. Change in Number of Requested Third-Party Domains and Cookies

	Requests		Cookies	
	(1) EU firm	(2) Non-EU firm	(3) EU firm	(4) Non-EU firm
Post × EU audience	−0.0789***	−0.0464***	−0.1096***	−0.0722***
	(0.0026)	(0.0028)	(0.0036)	(0.0042)
Post × non-EU audience	−0.0504***	−0.0220***	−0.0674***	−0.0310***
	(0.0038)	(0.0011)	(0.0051)	(0.0015)
Observations	961,587	2,691,711	961,587	2,691,711
$\overline{R^2}$	0.8871	0.9054	0.8444	0.8641
Mean dependent variable	2.3706	2.4647	1.3476	1.4296

	Request		Cookies	
	EU firm	Non-EU firm	EU firm	Non-EU firm
Post x EU audience	-0.0789 (0.0026)	-0.0464 (0.0028)	-0.1096 (0.0036)	-0.0722 (0.0042)
Post x non-EU audience	-0.0504 (0.0038)	-0.022 (0.0011)	-0.0674 (0.0051)	-0.0310 (0.0015)

Replication table 2

Table 5. Change in Number of Requested Third-Party Domains and Cookies: Website Popularity

	Requests		Cookies	
	(1) EU firm	(2) Non-EU firm	(3) EU firm	(4) Non-EU firm
Post × EU Audience	−0.0791***	−0.0473***	−0.1097***	−0.0734***
	(0.0026)	(0.0029)	(0.0036)	(0.0042)
Post × Non-EU Audience	−0.0505***	−0.0221***	−0.0676***	−0.0311***
	(0.0038)	(0.0011)	(0.0051)	(0.0015)
Post × EU Audience × Top1k	0.0312	0.0363**	−0.0027	0.0433*
	(0.0306)	(0.0160)	(0.0395)	(0.0232)
Post × Non-EU Audience × Top1k	−0.0010	0.0369	0.0885	0.0245
	(0.0727)	(0.0296)	(0.0943)	(0.0346)
Observations	960,993	2,689,302	960,993	2,689,302
$\overline{R^2}$	0.8870	0.9054	0.8444	0.8641
Mean dependent variable	2.3707	2.4651	1.3476	1.4300

	Request		Cookies	
	EU firm	Non-EU firm	EU firm	Non-EU firm
Post x EU audience	-0.0789 (0.0026)	-0.0464 (0.0028)	-0.1096 (0.0036)	-0.0722 (0.0042)
Post x non-EU audience	-0.0504 (0.0038)	-0.022 (0.0011)	-0.0674 (0.0051)	-0.0310 (0.0015)
Post x EU audience x Top1k	0.0312 (0.0306)	0.0363 (0.0369)	-0.0027 (0.0395)	0.0433 (0.0232)
Post x non-EU audience x Top1k	-0.0010 (0.0727)	0.0369 (0.0296)	0.0885 (0.0943)	0.0245 (0.0346)

Replication table 3

Table 6. Change in Websites Served by Data Type

	(1)	(2)	(3)
Post	0.013738	0.013287	0.015596
	(0.020129)	(0.020087)	(0.019912)
Post × *Collect Personal*	−0.056821**	−0.062199**	−0.066829**
	(0.026621)	(0.029627)	(0.029267)
Post × *Share Personal*		0.014936	−0.061624
		(0.036490)	(0.189001)
Post × *Collect Personal* × *Share Personal*			0.082506
			(0.192486)

	(1)	(2)	(3)
Post	0.0137 (0.0198)	0.0133 (0.0198)	0.0156 (0.0196)
Post x Collect Personal	-0.0568 (0.0262)	-0.0622 (0.0292)	-0.0668 (0.0288)
Post x Share Personal		0.0149 (0.0359)	-0.0616 (0.186)
Post x Collect Personal x Share Personal			0.0825 (0.1895)

Replication table 3

Table 7. Change in Market Structure of the Web Technology Industry

	(1) HHI all	(2) HHI without Google
After	16.016**	−0.291
	(6.517)	(0.606)
Observations	37	37
Pre-GDPR mean	950.833	70.090

	HHI all	HHI without Google
After	16.016 (6.517)	-0.291 (0.606)

Replication table 4

Table 8. Change in Levels and Market Shares: Google vs. All Other Firms

	(1) No. websites	(2) Market shares
Post × *Google*	−0.0261***	0.2731***
	(0.0000)	(0.0000)
Post × *Non-Google*	0.0006	−0.0000
	(0.0013)	(0.0000)
Observations	2,341,804	2,341,804
$\overline{R^2}$	0.8847	0.9994
Pre-GDPR Google	13.1370	29.6763
Pre-GDPR non-Google	0.6155	0.0011

	No. websites	Market shares
Post x Google	-0.0261 (0)	0.2731 (0)
Post x Google	0.0006 (0.0013)	-0 (0)

Replication table 5

Table 9. Change in Levels by Submarket

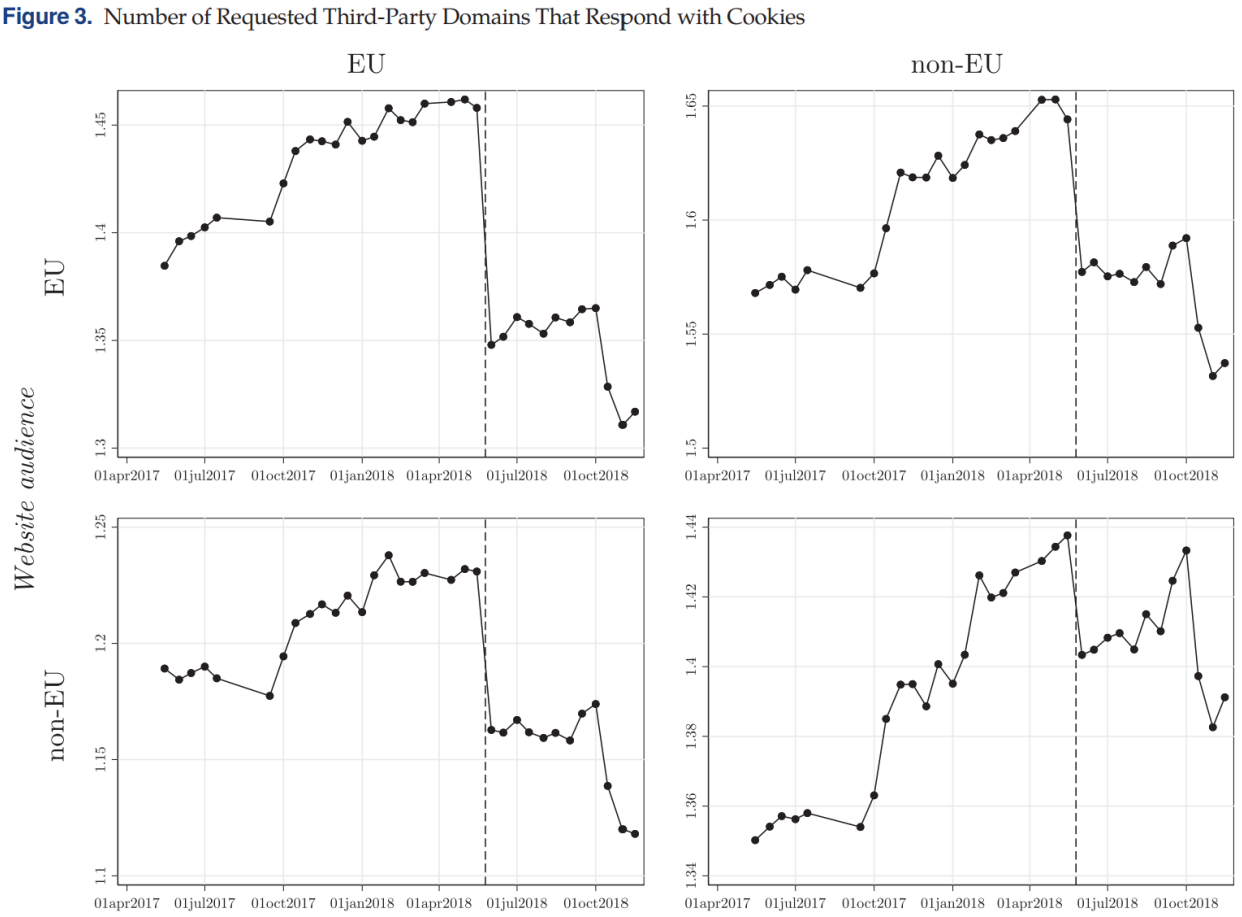
	(1) Video	(2) Advertising	(3) Analytics	(4) CDN/API	(5) Other
Post × Google	-0.012557***	-0.040471***	-0.01697***	-0.023483***	0.013068***
	(6.1e-18)	(1.4e-17)	(6.1e-17)	(4.1e-18)	(4.6e-17)
Post × Non-Google	-0.002325	-0.047413***	-0.026433***	-0.002923**	-0.006726
	(0.002143)	(0.014049)	(0.005668)	(0.001224)	(0.005318)
Observations	33,344	33,344	33,344	33,344	33,344
$\overline{R^2}$	0.986741	0.975713	0.983701	0.997838	0.982263
Pre-GDPR Google	10.1852	12.2644	11.4786	11.7228	6.69627
Pre-GDPR non-Google	0.1018	2.25412	0.921685	0.186973	0.40283

	Video	Advertising	Analytics	CDN/API	Others
Post x Google	-0.0126 (0)	-0.0405 (0)	-0.017 (0)	-0.0235 (0)	0.0131 (0)
Post x Non-Google	-0.0023 (0.0021)	-0.0474 (0.014)	-0.0264 (0.0057)	-0.0029 (0.0012)	-0.0067 (0.0053)

Extension

According to Figure 4 (Christian Peuker et, 2022), the authors examined the timeline from April 15th, 2017, to November 15th, 2018, which includes 19 months of data. With the GDPR event happening on May 25th, 2018, 13 months of data are used as before-event data, and six months of data are used as after-event data, which are not even lengthy. With a deep look at the data, we can find a slight tremendous increase from July 15th, 2017, to November 1st, 2017, which time sensitives may cause. Similar trends can be seen from September 15th, 2018, to October 1st, 2018. A tremendous increase may be caused by unspecified events encouraging audiences to visit websites more. For example, during summer vacation, students would have more time to visit websites, so there could be a significant increase during that period. We designed two extension experiments to question the above issues on estimating GDPR effects. First, to address the concern about time sensitivity, we would like to check specific periods from May 15th, 2017, to October 1st, 2017, and May 15th, 2018, to October 1st, 2018. In this case, we have May 15th, 2017, to October 1st, 2017, for the period before the GDPR event and May 15th, 2018, to October 1st, 2018, for the period after the GDPR event. By comparing the same period in different years, we can eliminate the effects of time sensitivity. Second, to get rid of the effects caused by different lengths before and after the GDPR event, we designed an experiment with a period from November 15th, 2017, to October 1st, 2018, which not only excludes the significant decrease since October 1st, 2018 but also keeps a relatively same time length before and after the events.

Figure 1: Figure from Origin Paper



Extension 1 - May - Oct 2017 V.S. May - Oct 2018

In extension 1, we did the same experiments as the original paper. Following are the results shown in Extension Table 1-6.

Discussion

Extension table 1 shows the change in number of requested third-party domains and cookies. Compared with Table 4 in the original paper, coefficients didn't change much, which means conclusions "web technologies in general are used less" and "web technologies with cookies are used less" are held.

Extension Table 1: corresponding to Table 4

	Requests	Cookies

	EU Firm	Non-EU Firm	EU Firm	Non-EU Firm
Post x EU audience	-.0689 (.0024)	-.0328 (.0025)	-.1107 (.0033)	-.0637 (.0037)
Post x non-EU audience	-.0387 (.0033)	-.0124 (.0009)	-.0655 (.0045)	-.0313 (.0013)
Observations	495363	1386639	495363	1386639
R2	.8781	.8982	.8340	.8550
Mean dependent variable	2.3547	2.4605	1.3281	1.4279

Extension table 2 shows the change in number of requested third-party domains and cookies based on website popularities. Compared with Table 5 in the original paper, most of the coefficients held a similar value. However, for non-EU audiences at top 1k websites, the requests for EU firms increased 0.128% whereas its original value -0.0010. Also for top 1k EU firms and EU audiences, instead of a minor decrease about 0.0027 in cookies percentage, now there are 0.048 increases.

Extension Table 2: corresponding to Table 5

	Requests		Cookies	
	EU Firm	Non-EU Firm	EU Firm	Non-EU Firm
Post x EU audience	-.0691(.0024)	-.0335 (.0026)	-.1109 (.0033)	-.0649 (.0038)
Post x non-EU audience	-.0388 (.0033)	-.0125 (.0009)	-.0656 (.0045)	-.0315 (.0013)
Post × EU Audience × Top1k	.0229 (.0345)	.0305 (.0190)	.0048 (.0449)	.0426 (.0273)
Post × Non-EU Audience × Top1k	.0128 (.0205)	.0557 (.0363)	.0518 (.1010)	.0564 (.0442)
Observations	495057	1385398	495057	1385398
R2	.8780	.8982	.8340	.8550
Mean dependent variable	2.3548	2.4608	1.3282	1.4282

Changes in market share of the web technology industry is shown in extension table 3. Typically they compared between all firms and firms without google. Compared with Table 7 in the original paper, we can see that the HHI (the sum of squared market shares) only increased by 10.063% from the base of 951.322 instead of 16%. Besides, Google's dominance becomes even more apparent, since HHI without Google decreased by 2.231% from a base of 71.486 instead of -0.291%.

Extension Table 3: Corresponding to Table 7

	HHI all	HHI without Google
After	10.063	-2.231
(std err)	14.244	2.239
Observations	20	20
Pre-GDPR mean	951.322	71.486

The changes in levels and market shares for Google and all other firms are shown in extension table 4. Compared with Table 8 in the original paper, we can see a smaller GDPR effect on Google and the effects on non google firms becomes negative.

Extension Table 4: Corresponding to Table 8

	No.websites	Market Shares
Post X Google	-0.0119***	0.2028***
(std err)	0.0	0.0
Post X Non-Google	-0.0004	-0.0
(std err)	0.0012	0.0
Observations	1265840	1265840
R2	0.8758	0.9993
Pre-GDPR Google	13.1191	29.6612
Pre-GDPR non-Google	13.1191	0.0011

Changes in levels by submarkets for Google and all other firms is shown in extension table 5. Submarkets, such as video, advertising, analytics, CDN/API, are discussed in extension table 11. We can see several differences compared with Table 9 in the original paper.First, google in video submarket has a weaker decrease compared with original paper whereas for non-google in video submarket, the effects of GDPR is smaller than original. Similar situations happened for subarea of advertising and analytics. Whereas in other aspects, the sign of the effects flapped for Google. In the original paper, the effects on Google in other submarkets is 0.013% increase, whereas, in extension 2, the effects on Google in other submarkets is -0.022%.

Extension Table 5: Corresponding to Table 9

	Video	Advertising	Analytics	CDN/API	Other
Post X Google	-0.0036***	-0.0214***	-0.0145***	-0.0074***	-0.022***
(std err)	0.0	0.0	0.0	0.0	0.0
Post X Non-Google	-0.0014	-0.0441***	-0.0172***	-0.0017	-0.0032
(std err)	0.0027	0.0111	0.0046	0.0011	0.0043
Observations	17714	17714	17714	17714	17714
R2	0.9843	0.9723	0.9808	0.9974	0.9798
Pre-GDPR Google	10.177214	12.235809	11.47565	11.718884	6.683246
Pre-GDPR non-Google	0.102427	2.242642	0.910283	0.184549	0.399961

Next extension table 12 discussed the change in market shares among different submarkets. There are several changes in the effects. First, Google in the video submarket has a weaker decrease than the original paper. Similar situations happened for subarea of advertising and analytics.Whereas in other aspects, the sign of the effects flapped again for Google. In the original paper, the effects on Google in other submarkets is 0.025% increase, whereas, in extension 2, there was almost no effect on Google.

Extension Table 6: Corresponding to Table 10

	Video	Advertising	Analytics	CDN/API	Other
Post X Google	-0.1586***	0.3274***	0.9137***	-0.3753***	-8.5e-05***
(std err)	0.0	0.0	0.0	0.0	0.0
Post X Non-Google	0.0002	-0.0003	-0.0009	0.0004	0.0
(std err)	0.0008	0.0007	0.0007	0.0007	0.0018
Observations	17714	17714	17714	17714	17714
R2	0.9997	0.9985	0.9976	0.9998	0.9791
Pre-GDPR Google	74.374786	26.848364	39.871235	71.293312	1.209089
Pre-GDPR non-Google	0.024616	0.070271	0.057761	0.027576	0.0949

Extension 2: Nov 2017 - Oct 2018

In extension 2, we did the same experiments as the original paper. Following are the results shown in Extension Table 7-12.

Discussion

Extension table 7 shows the change in number of requested third-party domains and cookies. Compared with Table 4 in the original paper, coefficients didn't change much, which means conclusions "web technologies in general are used less" and "web technologies with cookies are used less" are held.

Extension Table 7: Corresponding to Table 4

	Requests		Cookies	
	EU Firm	Non-EU Firm	EU Firm	Non-EU Firm
Post x EU audience	-.0693 (.0024)	-.0395 (.0025)	-.1133 (.0033)	-.0804 (.0037)
Post x non-EU audience	-.0401 (.0034)	-.0150 (.0009)	-.0747 (.0046)	-.0446 (.0013)
Observations	611919	1712907	611919	1712907
R2	.9076	.9280	.8741	.8974
Mean dependent variable	2.3778	2.4773	1.356	1.4513

Extension table 8 shows the change in number of requested third-party domains and cookies based on website popularities. Compared with Table 6 in the original paper, most coefficients held a similar value. However, for non-EU audiences at top 1k websites, the requests for EU firms significantly decreased to -0.454 compared with its original value -0.0010. Also for top 1k EU firms and EU audiences, instead of a minor decrease about 0.0027 in cookies percentage, now there are 0.0360 increases.

Extension Table 8: Corresponding to Table 5

	Requests		Cookies	
	EU Firm	Non-EU Firm	EU Firm	Non-EU Firm
Post x EU audience	-.0696(.0024)	-.0405 (.0026)	-.1136 (.0033)	-.0817 (.0038)
Post x non-EU audience	-.0402 (.0034)	-.0151 (.0009)	-.0748 (.0046)	-.0447 (.0013)
Post × EU Audience × Top1k	.0539 (.0276)	.0381 (.0137)	.0360 (.0371)	.0470 (.0201)
Post × Non-EU Audience × Top1k	-.0454 (.1193)	.0334 (.0252)	.0452 (.1293)	.0277 (.0299)
Observations	611541	1711374	611541	1711374
R2	.9075	.9280	.8741	.8974
Mean dependent variable	2.3779	2.4777	1.3564	1.4517

Changes in market share of the web technology industry is shown in extension table 9. Typically they compared between all firms and firms without google. Compared with Table 7 in the original paper, we can see that the HHI (the sum of squared market shares) only increased by 8.176% from the base of 950.82 instead of 16%. Besides, Google's dominance becomes even more

apparent, since HHI without Google decreased by 1.576% from a base of 68.458 instead of -0.291%.

Extension Table 9: Corresponding to Table 7

	HHI all	HHI without Google
After	8.176	-1.576
(std err)	9.221	1.129
Observations	21	21
Pre-GDPR mean	950.82	68.458

The changes in levels and market shares for Google and all other firms are shown in extension table 10. Compared with Table 8 in the original paper, we can see a smaller GDPR effect on Google.

Extension Table 10: Corresponding to Table 8

	No.websites	Market Shares
Post X Google	-0.0192***	0.161***
(std err)	0.0	0.0
Post X Non-Google	0.0013	-0.0
(std err)	0.0011	0.0
Observations	1329132	1329132
R2	0.9232	0.9996
Pre-GDPR Google	13.1587	29.7036
Pre-GDPR non-Google	13.1587	0.0011

Changes in levels by submarkets for Google and all other firms is shown in extension table 11. Submarkets, such as video, advertising, analytics, CDN/API, are discussed in extension table 11. We can see several differences compared with Table 9 in the original paper.First, google in video submarket has a stronger decrease compared with original paper whereas for non-google in video submarket, the effects of GDPR is smaller than original. Similar situations happened for subarea of advertising. In analytics, non-google firms experienced a smaller effect than the original paper. Whereas in other aspects, the sign of the effects flapped for Google. In the original paper, the effects on Google in other submarkets is 0.013% increase, whereas, in extension 2, the effects on Google in other submarkets is -0.0057%.

Extension Table 11: Corresponding to Table 9

	Video	Advertising	Analytics	CDN/API	Other
Post X Google	-0.0312***	-0.0265***	-0.0146***	-0.0105***	-0.0057***
(std err)	0.0	0.0	0.0	0.0	0.0
Post X Non-Google	-0.0006	-0.0233**	-0.0099**	-0.0012	-0.006
(std err)	0.0021	0.0111	0.0049	0.0008	0.0042
Observations	20840	20840	20840	20840	20840
R2	0.9889	0.984	0.9893	0.9987	0.9864
Pre-GDPR Google	10.197235	12.287216	11.480545	11.726396	6.707033
Pre-GDPR non-Google	0.101116	2.257271	0.928712	0.188791	0.405099

Next extension table 12 discussed the change in market shares among different submarkets. There are several changes in the effects. First, Google in the video submarket has a stronger decrease than the original paper. Similar situations happened for subarea of advertising. In analytics, Google has a smaller market share increase compared with the original paper . Whereas in other aspects, the sign of the effects flapped again for Google. In the original paper, the effects on Google in other submarkets is 0.025% increase, whereas, in extension 2, the effects on Google in other submarkets is -0.029%.

Extension Table 12: Corresponding to Table 10

	Video	Advertising	Analytics	CDN/API	Other
Post X Google	-1.6969***	0.4869***	0.4948***	-0.5692***	-0.029***
(std err)	0.0	0.0	0.0	0.0	0.0
Post X Non-Google	0.0016	-0.0005	-0.0005	0.0005	2.8e-05
(std err)	0.001	0.0007	0.0007	0.0006	0.0016
Observations	20840	20840	20840	20840	20840
R2	0.9997	0.9991	0.9986	0.9999	0.9954
Pre-GDPR Google	77.04818	27.228949	39.232204	70.080956	1.192459
Pre-GDPR non-Google	0.022048	0.069905	0.058374	0.028741	0.094916

Conclusion

With extension 1 and extension 2, we can see that most of the conclusions about EU audience, non-EU audience and EU firm, Non-EU firm are held. However, with extension 1, where the examined time is the same months in different years, effects of GDPR tend to be weaker but still

significant. And in extension 2, we can still see a significant effect on firms and audience from the GDPR events, but sub-areas such as videos and others have stronger effects than before. Overall, we think that the GDPR effect is significant.

Bibliography

Christian Peukert, Stefan Bechtold, Michail Batikas, Tobias Kretschmer (2022) Regulatory Spillovers and Data Governance: Evidence from the GDPR. Marketing Science 41(4):746-768. https://doi.org/10.1287/mksc.2021.1339