

Privacy & market concentration: Intended & unintended consequences of the GDPR

Garrett Johnson (Questrom, Boston U)

Scott Shriver (Leeds, U Colorado Boulder)

Sam Goldberg (Kellogg, Northwestern U)

Can privacy policy hurt competition?

Economies of scale

Larger firms have more resources for compliance

B2B choice of data vendors

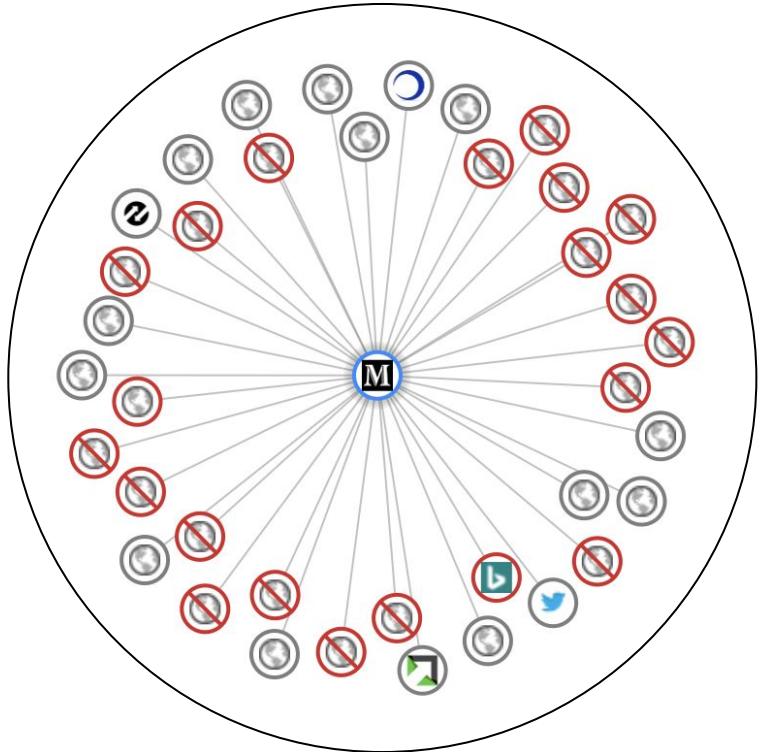
Firms may prefer larger vendors because they have better products

Overview: GDPR & Web technology

- General Data Protection Regulation (GDPR)
 - Enforced as of May 25, 2018
 - ↑ **logistical cost & legal risk of personal data processing**
- **Technology vendors** provide an **ecosystem** for the web to **thrive**
 - Vendors help websites: monetize with ads, load & share content, measure & optimize site traffic
 - Vendors often share personal data
 - Regulatory **scrutiny**, but no fines to date

Overview

- Data & theory
- Results
 - GDPR effect on vendor use
 - Concentration of vendor market
 - Heterogeneity by website



Data & theory

Advertisement

DECATHLON



JE DÉCOUVRE >

Support The Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

Search jobs



Sign in



Search

International
edition ▾

The Guardian

News

Opinion

Sport

Culture

Lifestyle

More ▾

Books Music TV & radio Art & design Film Games Classical Stage

Books

The Great Gatsby prequel set for release days after copyright expires

Michael Farris Smith's book **Nick** will tell story of Nick Carraway's life before meeting the millionaire of F Scott Fitzgerald's novel



Alison Flood

Wed 15 Jul 2020 14.47
BST

GDPR Impact



15



Gagnez votre
déménagement avec
Bouygues Telecom.

Advertisement by Bouygues Telecom

Advertisement

DECATHLON

Panier de
TARM
139

Support The Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

News

Opinion

Sport

Books Music TV & radio Art & design Film Games

Books

The Great Gatsby days after copy

Michael Farris Smith's book **Nick** will tell story of Nick Carraway's life before meeting the millionaire of F Scott Fitzgerald's novel

Alison Flood

Wed 15 Jul 2020 14.47

BST

GDPR Impact



DISCONNECT

Browse the web normally. As you do, the graph in this popup and the counter in the toolbar will update. Each circle in the graph represents a site that's been or would've been sent some of your personal info.

Circles with a halo are sites you've visited. Circles without a halo are sites you haven't.

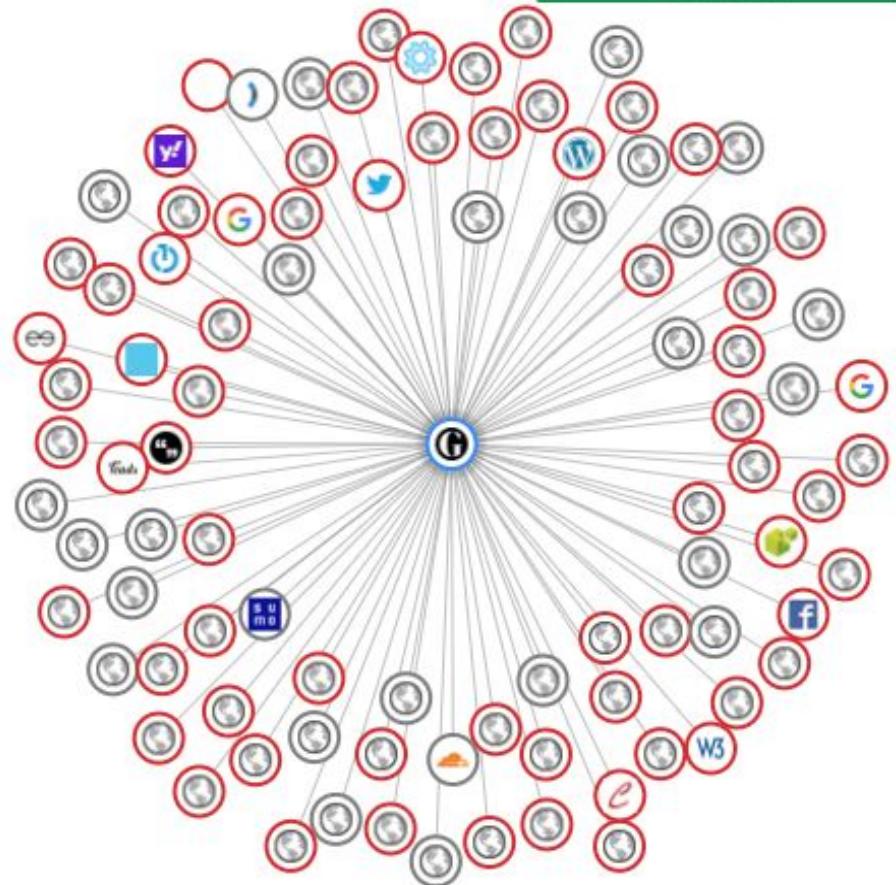
Red circles are known tracking sites. Gray circles aren't but may still track you.

Mouse over a circle to view that site's tracking footprint. Click a red circle to block or unblock that site.

Block tracking sites

Hide sidebar

Show list view



Collect data 3rd party domains on top websites

1. Use VPN service to simulate EU user (France)
2. Use "webxray" program (Libert 2015) to record all 3rd party domains when visiting a website
3. Repeat for 28K top sites regularly in 2018 (14 times)
 - Top 2,000 websites in
 - Each of 28 EU countries
 - US
 - Canada
 - Globally

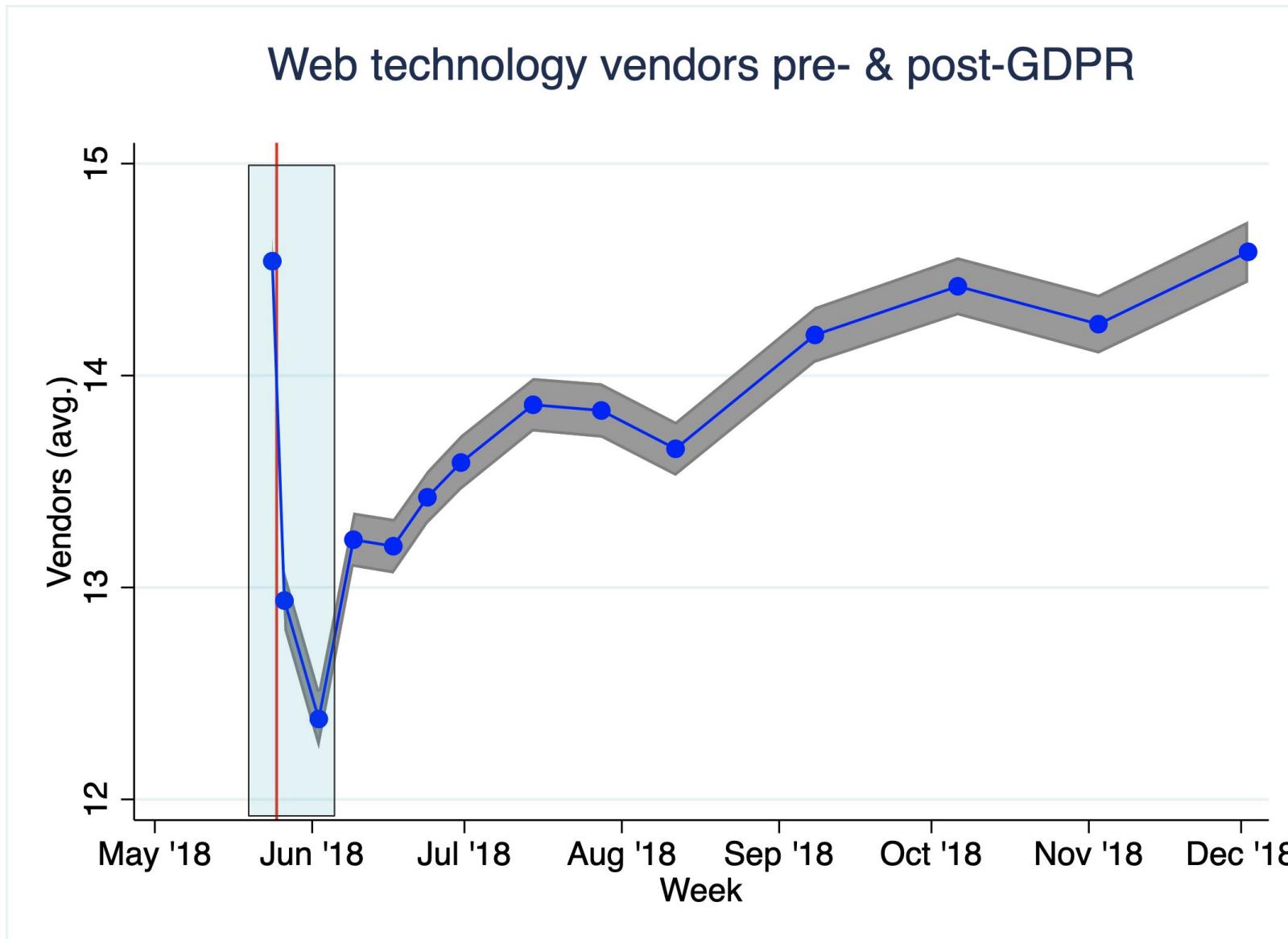
Theory: Website vendor choices under GDPR

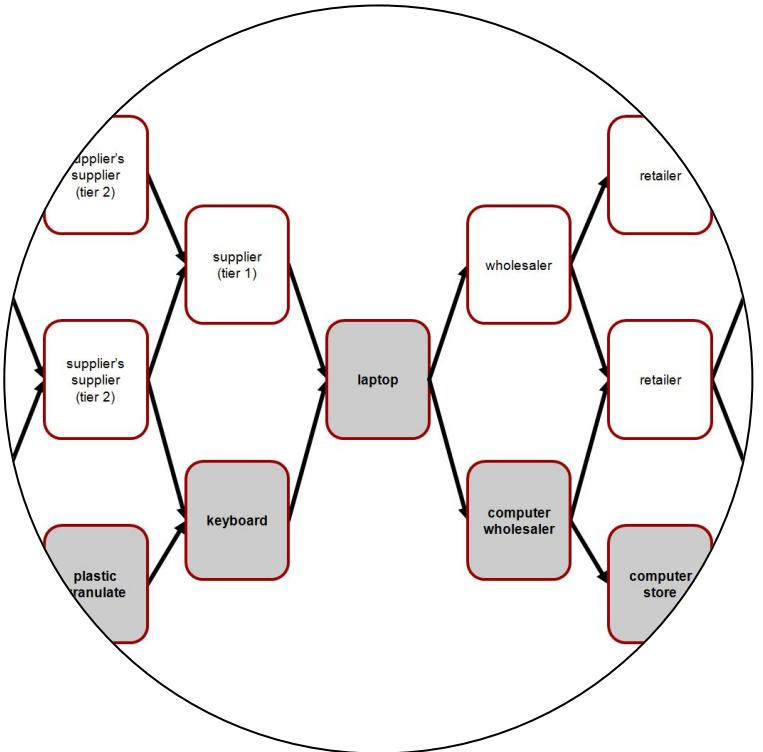
- Equilibrium model of web technology industry
 - Differentiated vendors with asymmetric costs
 - Representative website chooses vendors that improve revenue
- Regulatory incentive: $\text{Fine} * \text{Pr}[\text{Fine}]$
 - $\text{Fine} = 4\%$ of global revenue
 - $\text{Pr}[\text{Fine}]$: Sites beliefs about enforcement
- **Key comparative statics:** As site's GDPR enforcement beliefs ↑,
 - ↓ active vendors in equilibrium
 - ↑ concentration in vendor market
 - High-revenue sites cut vendors the most



Results: GDPR effect on vendor use

Short run: vendors fall 15% 1 week post GDPR

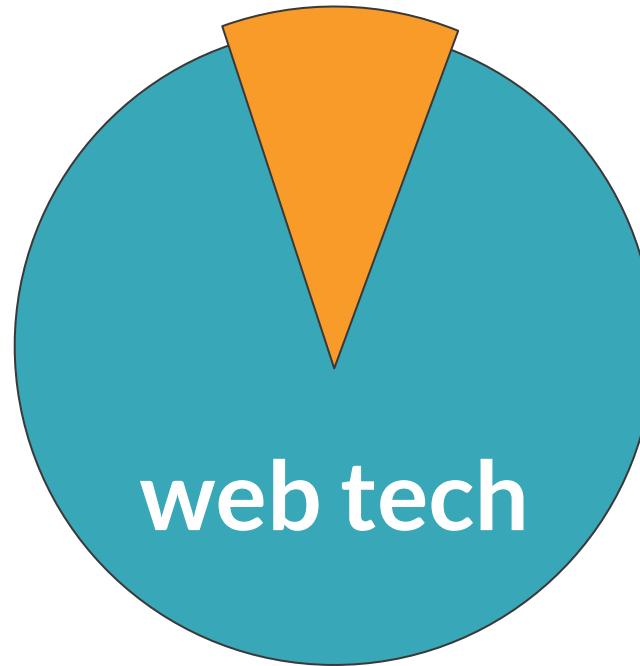




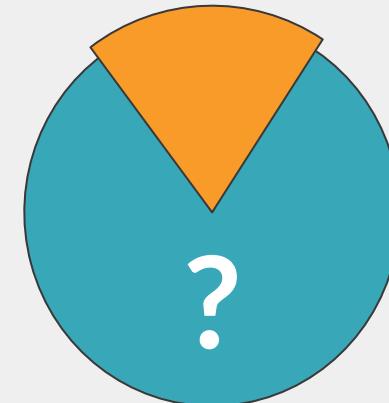
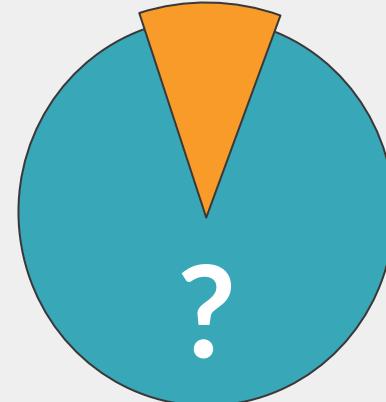
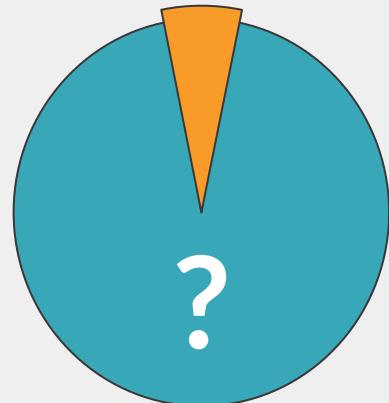
Results: Concentration of vendor market

Do larger vendors get a larger share of the smaller pie?

Pre-GDPR

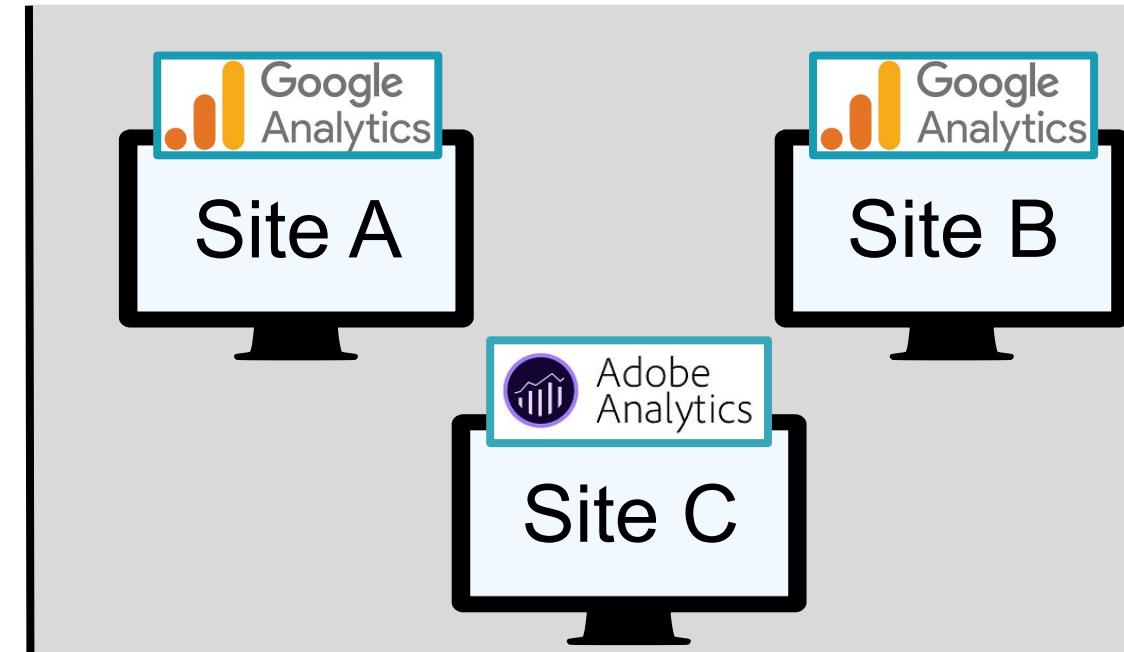


Post-GDPR



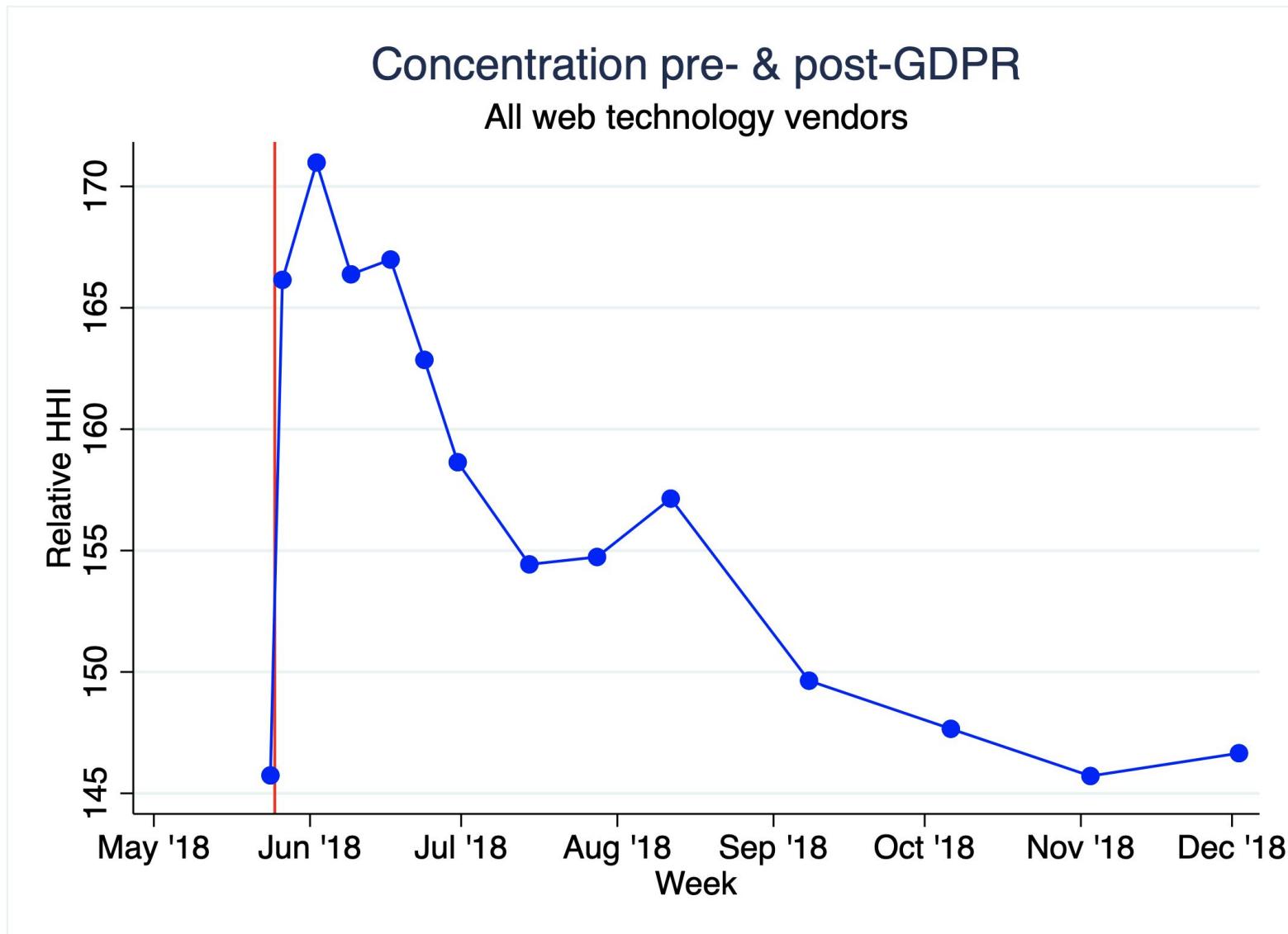
Defining relative market concentration

- **Reach:** # of websites using vendor
- **Vendor's relative market share:**
$$\text{market share} = \text{own-reach} / \sum \text{reach}$$
- **Herfindahl-Hirschman Index (HHI)**
$$\sum \text{market share}^2$$



	Reach	Share
Google Analytics	2	2/3
Adobe Analytics	1	1/3

Short run: concentration rises 17% post-GDPR



↑ concentration in top 4 categories (94.3% of vendors)

Category	Vendor Share (Pre)	Pre	HHI Post	Diff. (%)
All vendors		146	171	17.3%
All categorized vendors		308	363	17.8%
Advertising	50.2%	348	436	25.3%
Hosting	20.5%	1,892	1,936	2.3%
Audience measurement	14.4%	4,116	4,355	5.8%
Social media	9.2%	4,251	4,412	3.8%

Extension: Personal data concentrated in top vendors

Data samples	HHI Pre	HHI Post	Diff.	Diff. (%)
<i>Role of personal data</i>				
Likely personal data	185.9	221.7	35.8	19.3%
Unlikely personal data	487.4	515	27.5	5.6%

Extension: Consent does not drive ↑ concentration

Less surprising because most consent popovers bury vendor list

Data samples	HHI Pre	HHI Post	Diff.	Diff. (%)
<i>Role of personal data</i>				
Likely personal data	185.9	221.7	35.8	19.3%
Unlikely personal data	487.4	515	27.5	5.6%
<i>Role of consent</i>				
Sites using consent platform	100.1	117.9	17.8	17.8%
Sites without consent platform	153.6	179.4	25.8	16.8%

Extension: Google & Facebook drive ↑ concentration

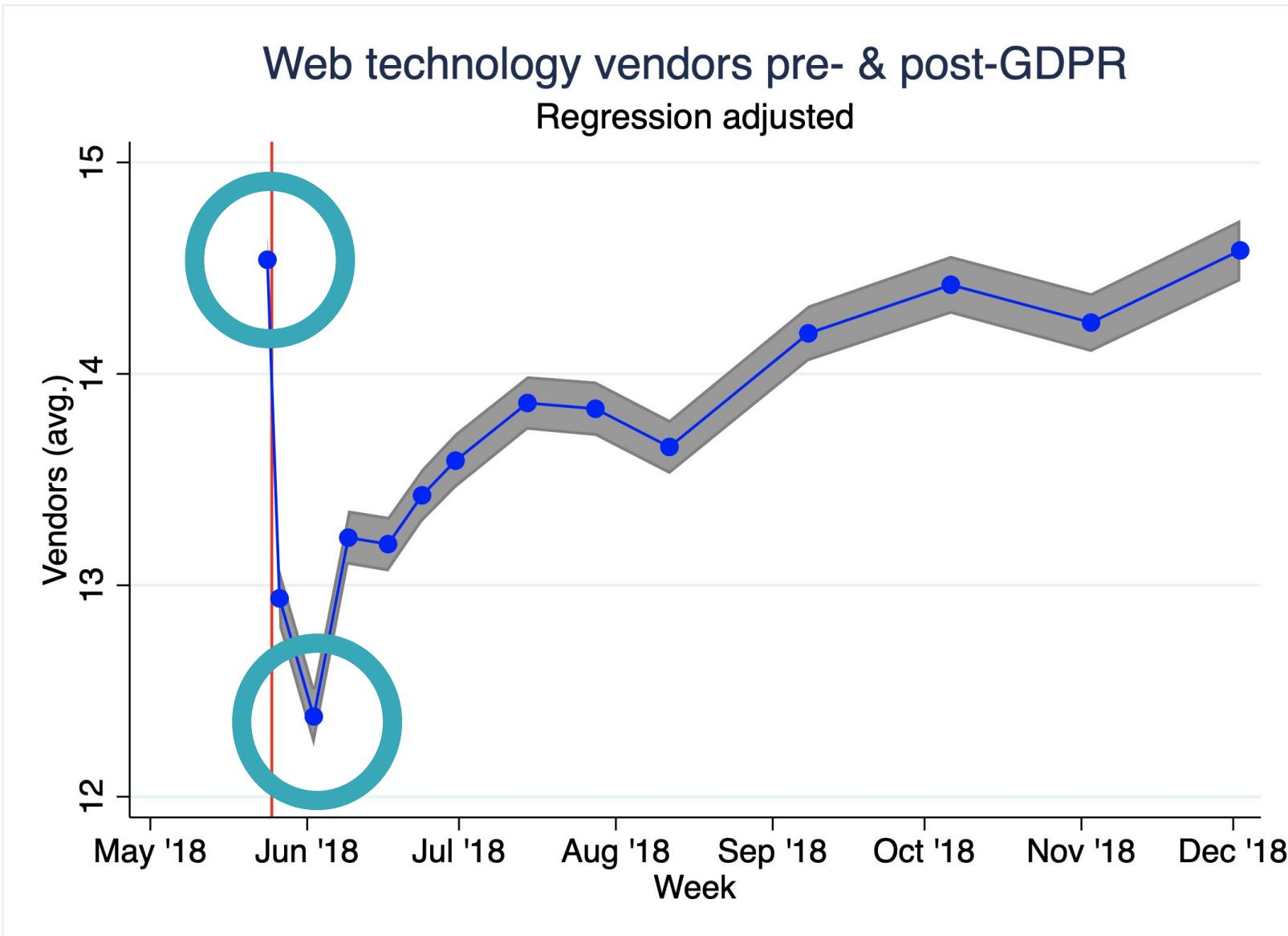
Data samples	HHI Pre	HHI Post	Diff.	Diff. (%)
<i>Role of personal data</i>				
Likely personal data	185.9	221.7	35.8	19.3%
Unlikely personal data	487.4	515	27.5	5.6%
<i>Role of consent</i>				
Sites with privacy extension	100.1	117.9	17.8	17.8%
Sites without privacy extension	153.6	179.4	25.8	16.8%
<i>Role of top 2 companies (Google & Facebook)</i>				
All vendors	145.7	171.0	25.2	17.3%
All but top 2 companies	46.0	43.2	-2.8	-6.2%



Results: Heterogeneity by website

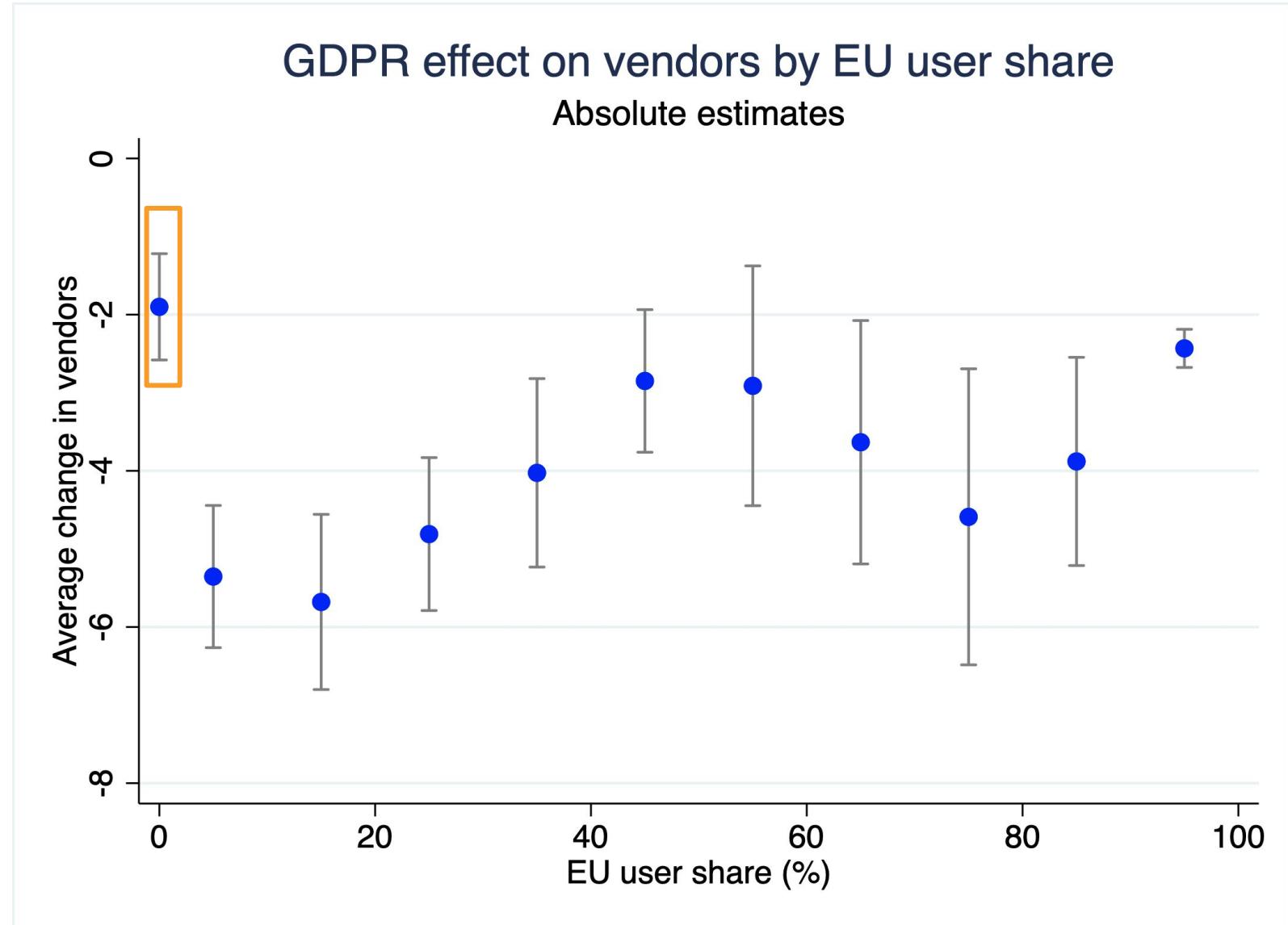
Heterogeneity in short run

Short run = full week after enforcement deadline vs pre-deadline baseline



Fine on global revenues most deters non-EU sites

Discontinuity for sites with 0% EU users: these sites not subject to the GDPR



Dependent variable	Vendors
GDPR	-3.850*** (0.382)
GDPR X Share of EU users (%)	0.0128*** (0.003)
GDPR X No EU users	2.233*** (0.301)
GDPR X log(Site rank)	0.209*** (0.042)
GDPR X log(Ad count + 1)	-1.541*** (0.080)
GDPR X User income	-0.363*** (0.057)
Site fixed effects	X
Observations	52,095
R-squared	0.903
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

GDPR incentives explain site cuts to vendors

Dependent variable	Vendors
GDPR	-3.850*** (0.382)
GDPR X Share of EU users (%)	0.0128*** (0.003)
GDPR X No EU users	2.233*** (0.301)
GDPR X log(Site rank)	0.209*** (0.042)
GDPR X log(Ad count + 1)	-1.541*** (0.080)
GDPR X User income	-0.363*** (0.057)
Site fixed effects	X
Observations	52,095
R-squared	0.903
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

GDPR incentives explain site cuts to vendors

GDPR's global fines especially deter non-EU sites

Dependent variable	Vendors
GDPR	-3.850*** (0.382)
GDPR X Share of EU users (%)	0.0128*** (0.003)
GDPR X No EU users	2.233*** (0.301)
GDPR X log(Site rank)	0.209*** (0.042)
GDPR X log(Ad count + 1)	-1.541*** (0.080)
GDPR X User income	-0.363*** (0.057)
Site fixed effects	X
Observations	52,095
R-squared	0.903
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

GDPR incentives explain site cuts to vendors

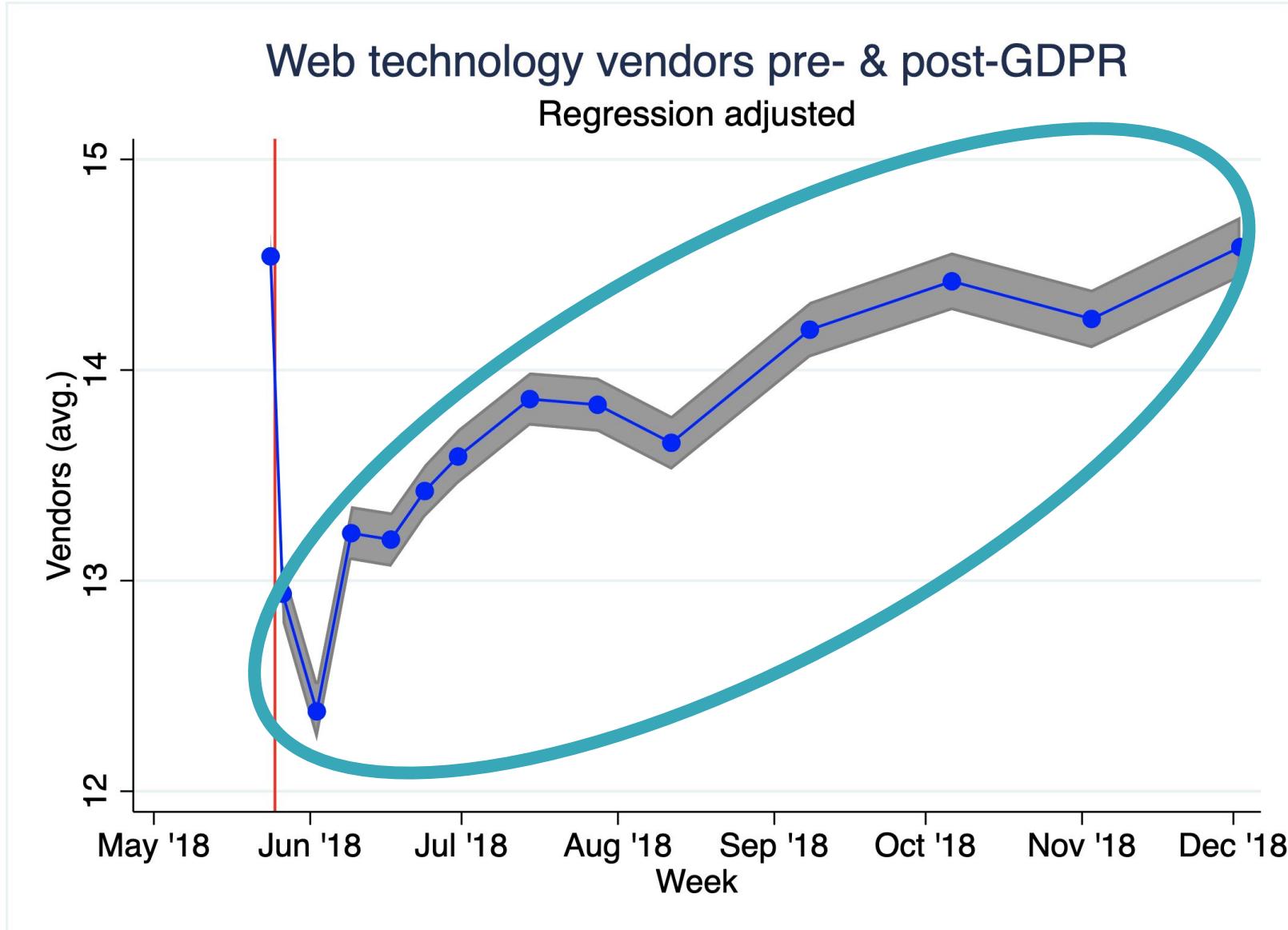
GDPR's global fines especially deter non-EU sites

Sites with high revenue sites cut vendors most in short run

- Traffic (inverse of site rank)
- Ad count
- User income

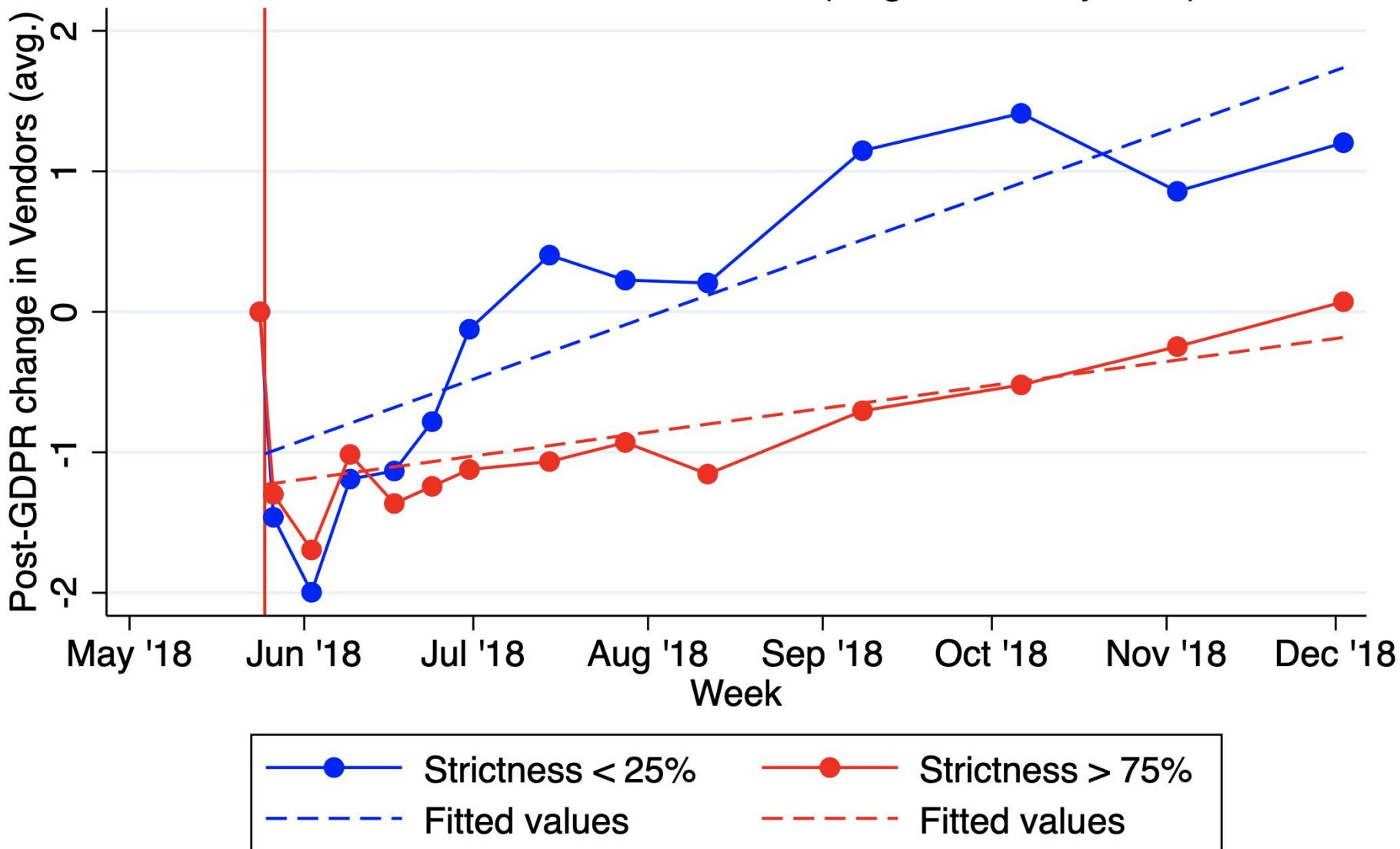
Heterogeneity in long run evolution

Long run evolution = post-GDPR trend in 2018



Vendor evolution by regulatory strictness

Publisher with > 50% EU users (Regression adjusted)



Dependent variable [†]	Vendors
Post-GDPR trend	0.0352* (0.0211)
Trend X Share of EU users (%)	0.000373* (0.0002)
Trend X Regulatory strictness	-0.00359* (0.0019)
Trend X log(Site rank)	-0.00600*** (0.0014)
Trend X log(Ad count + 1)	0.0678*** (0.0024)
Trend X User income	-0.0104*** (0.0015)
Site fixed effects	X
Observations	240,545
R-squared	0.837
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Post-GDPR vendor rise seems related to enforcement beliefs

†EU site sample: $\geq 50\%$ EU user traffic

Dependent variable [†]	Vendors
Post-GDPR trend	0.0352* (0.0211)
Trend X Share of EU users (%)	0.000373* (0.0002)
Trend X Regulatory strictness	-0.00359* (0.0019)
Trend X log(Site rank)	-0.00600*** (0.0014)
Trend X log(Ad count + 1)	0.0678*** (0.0024)
Trend X User income	-0.0104*** (0.0015)
Site fixed effects	X
Observations	240,545
R-squared	0.837
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Post-GDPR vendor rise seems related to enforcement beliefs

Websites facing stricter local data regulator keep vendor use low

†EU site sample: $\geq 50\%$ EU user traffic

Dependent variable [†]	Vendors
Post-GDPR trend	0.0352* (0.0211)
Trend X Share of EU users (%)	0.000373* (0.0002)
Trend X Regulatory strictness	-0.00359* (0.0019)
Trend X log(Site rank)	-0.00600*** (0.0014)
Trend X log(Ad count + 1)	0.0678*** (0.0024)
Trend X User income	-0.0104*** (0.0015)
Site fixed effects	X
Observations	240,545
R-squared	0.837
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Post-GDPR vendor rise seems related to enforcement beliefs

Websites facing stricter local data regulator keep vendor use low

High revenue sites increase vendors most post-GDPR

- Traffic (inverse of site rank)
- Ad count
- ~~User income~~

[†]EU site sample: ≥50% EU user traffic

Summary: GDPR's intended & unintended consequences

- Novel empirical evidence of **privacy-competition policy tension**
- **Intended consequence:** ↓ web tech data sharing
- **Unintended consequences:**
 - ↑ vendor concentration
 - Sites with most EU visitors reduce vendors the least



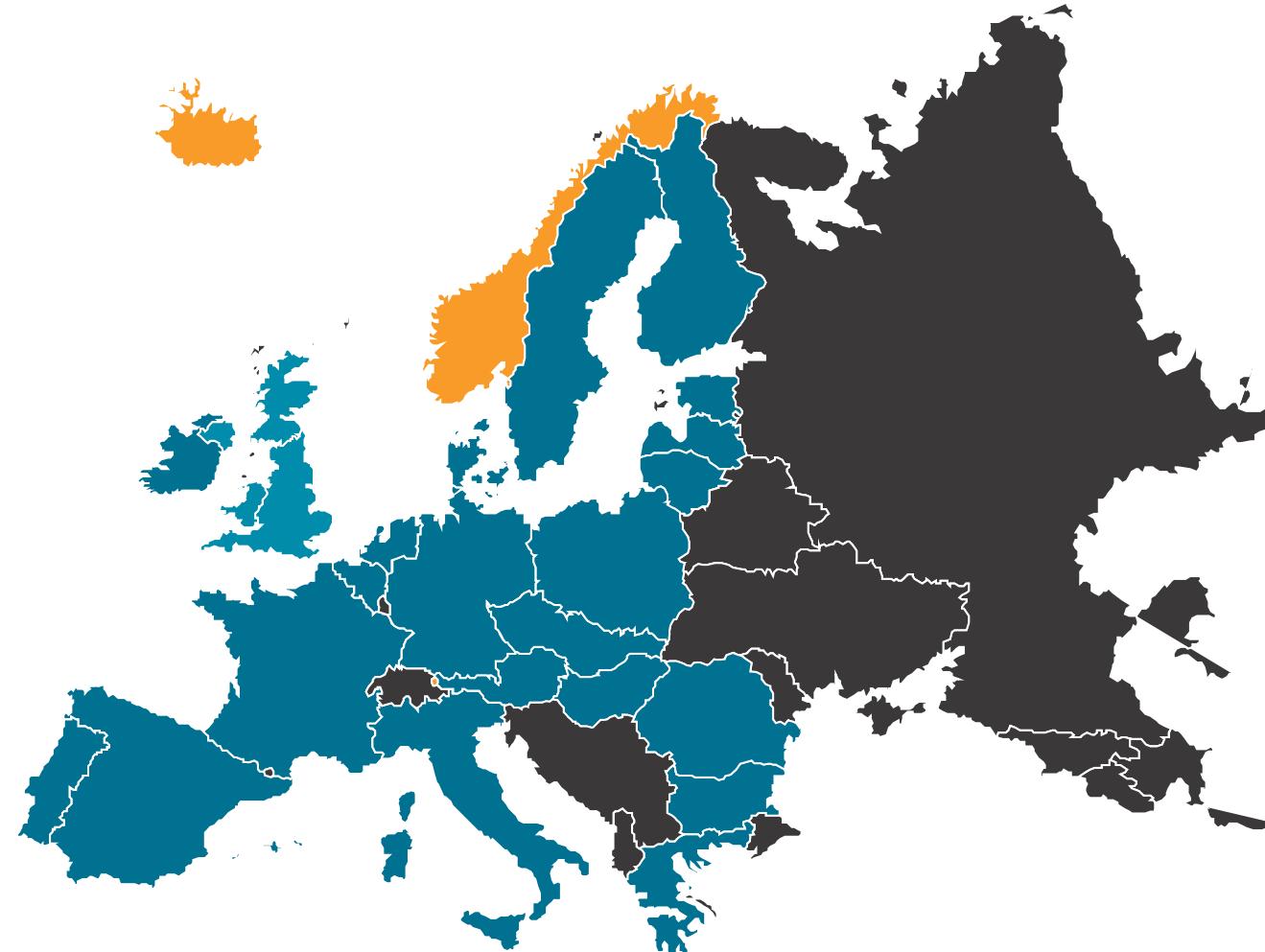
THE CONCENTRATION IS

TOO DAMN HHI

Appendix

GDPR

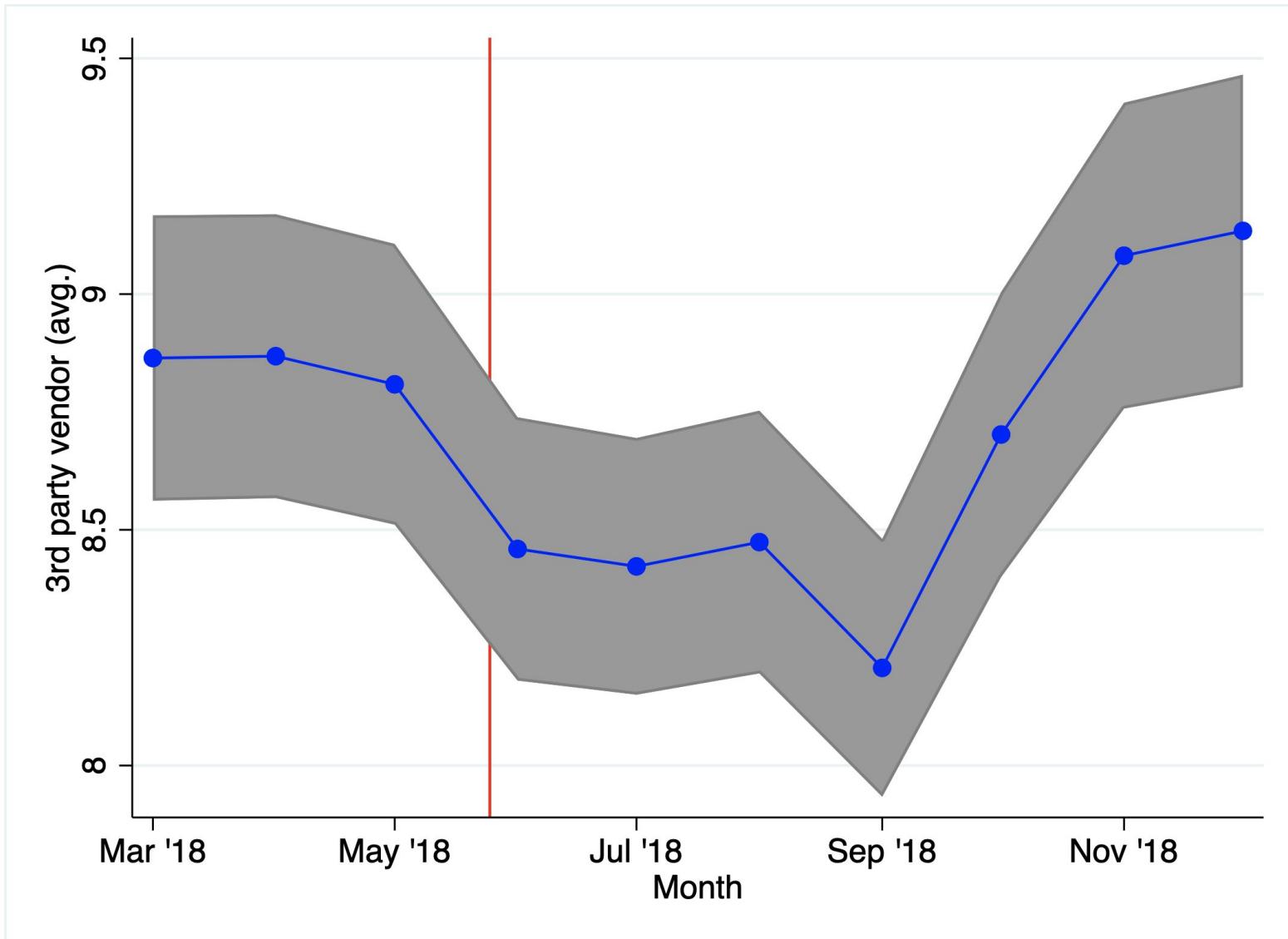
General Data Protection Regulation



EU
EEA
Brexit

No pre-GDPR trend in vendors

1,452 site sample overlapping with ours from WhoTracksMe dataset



Fewer vendors in all categories but compliance

Short run estimates: 1 week post-GDPR

Category	Avg. Vendors		
	Pre	Post	Diff. (%)
All vendors	14.44	12.35	-14.5%
All categorized vendors	8.40	6.91	-17.7%
Advertising	4.35	3.29	-24.3%
Hosting	1.78	1.61	-9.7%
Audience measurement	1.25	1.11	-10.9%
Social media	0.79	0.70	-11.5%
Design optimization	0.22	0.20	-10.5%
Security	0.15	0.12	-17.8%
Native ads	0.078	0.066	-14.8%
CRM	0.022	0.019	-9.6%
Privacy compliance	0.017	0.021	23.2%

↑ concentration in top 4 categories (94.3% of vendors)

Category	Vendor Share (Pre)	HHI		
		Pre	Post	Diff. (%)
All vendors		146	171	17.3%
All categorized vendors		308	363	17.8%
Advertising	50.2%	348	436	25.3%
Hosting	20.5%	1,892	1,936	2.3%
Audience measurement	14.4%	4,116	4,355	5.8%
Social media	9.2%	4,251	4,412	3.8%
Design optimization	2.6%	2,874	2,861	-0.5%
Security	1.8%	8,926	9,722	8.9%
Native ads	0.9%	4,229	4,024	-4.8%
CRM	0.2%	6,408	6,119	-4.5%
Privacy compliance	0.2%	3,925	4,116	4.9%

Robustness of concentration effect

- Defining markets by vendor category
 - Categories include: advertising, social media, site analytics
 - Two categorizations
 - *Webxray*: ↑ Concentration in top 4 categories (94% of total)
 - *WhoTracksMe*: ↑ Concentration in top 3 categories (85% of total)
- Alternate concentration metrics
 - Concentration ratios: Total share of top N firms
 - "Head-to-head competition": Probability site retains dominant firm given drops only one of two vendors