

Taxing Multinational Digital Platforms: Evidence from France’s Digital Services Tax

Gerardo Sanz-Maldonado*

Department of Economics, University of Michigan

sanzg@umich.edu

Draft: October 19, 2026

[Click Here For Latest Version](#)

Abstract

Digital service taxes are levies on the revenues of large multinational digital platforms attributed to local activity. This paper shows that their design generates behavioral incentives that can change the pattern of web traffic among liable sites and their non-liable counterparts. I study France’s digital services tax using monthly traffic data for 55 website families owned by 20 liable corporate groups. I estimate an imprecise average traffic decline of 2.8 log percentage points among liable platforms. Further, the point estimate for nonliable sites exposed to digital advertising technologies is a decline of 8.2 log percentage points. Treatment effects among liable platforms are heterogeneous by business model: pure intermediation websites decline by 11.1 log percentage points, while advertising websites increase by 16.9 log percentage points. A model explains how reduced advertising intensity can attract users despite lower expenditure on services, while higher transaction fees on intermediation platforms discourage participation. The findings extend the assessment of digital taxation beyond price incidence.

1 Introduction

Digital services taxes are gross turnover taxes on the local revenues of large multinational digital platforms. In response to these taxes, platforms may change the services they provide, the advertising users encounter, or the terms on which buyers and sellers transact. Importantly, these adjustments can affect users even when they pay no monetary price to access a platform.

Many of these adjustments are difficult to observe directly. Public financial statements rarely reveal how platforms allocate expenditure across services and geographic markets. Moreover, the platforms subject

*I am grateful to my committee—Jim Hines, Joel Slemrod, Matthew Shapiro, and Damian Vergara—for their guidance and support. I also thank discussants at the University of Michigan PF Free Lunch Seminar and the 2026 International Conference of the IIPF for helpful comments and suggestions.

to DSTs often operate as “walled gardens”: they control access to their services and the data generated by users, limiting what outside researchers can observe. These restrictions, together with stark differences in business models and pricing strategies, complicate the search for a common yardstick with which to compare platforms’ responses to taxation. This paper’s central methodological point is that web traffic measures provide such a yardstick, allowing comparisons across countries, platforms, and over time.

The key observation is that a user’s decision to visit a platform reflects the value that platforms provide to end-users. This value, in turn, depends on the varied inputs into digital production—from content provision and advertising intensity to the terms on which buyers and sellers transact. Changes in traffic therefore provide evidence about how users respond to the platform choices that digital services taxes may alter.

I study these responses around the December 2018 announcement of France’s 3 percent digital services tax on targeted advertising and digital intermediation. Using monthly website-popularity data, I estimate that French traffic to 55 website families owned by 20 liable corporate groups increases by 2.6 log percentage points (s.e. 6.76) over January–December 2019 relative to synthetic controls constructed from the full pool of nonliable websites by fitting prepolicy French traffic histories. This near-zero estimate is imprecise, and the point estimate varies with advertising exposure among comparison sites. In the preferred comparison, synthetic controls are drawn from nonliable sites without observed connections to liable-owned WTM advertising technologies in French browsing during September–November 2018. Relative to these controls, traffic to liable sites declines by 2.8 log percentage points (s.e. 7.78), while traffic to advertising-exposed nonliable sites declines by 8.2 log percentage points (s.e. 4.96).

It’s important to recognize that the design of the tax targets two tax bases that interact with user value in fundamentally different ways. While a turnover tax tends to depress the return to taxed activities and generally result in less, for example, digital advertising and e-commerce intermediation, the impact of these changes on user traffic push in opposite directions. For an advertising-financed platform, user-facing expenditure attracts attention, while advertising monetizes that attention but can discourage visits. A tax on advertising revenue reduces the return to providing content and services, but it can also induce the platform to reduce advertising intensity. The resulting traffic response depends on which adjustment dominates. For a platform that intermediates transactions, higher fees instead raise buyer prices and reduce seller supply, making the platform less attractive to users through both channels. Later on, I develop a pair of models that formalize this intuition.

To test this model, I measure advertising’s share of advertising plus intermediation revenue, using publicly available parent-company consolidated accounting records for fiscal year 2018. Because this measure is defined at the consolidated level, it serves as a proxy for each site’s underlying business model.

I estimate a heterogeneous difference-in-differences model for the 55 liable website families, using synthetic controls constructed from nonliable sites without documented advertising-tracker exposure. I standardize advertising’s share of in-scope revenue across the liable sample. A one-standard-deviation increase in this share—36.4 percentage points—is associated with a 12.38 log percentage point increase in traffic (s.e. 4.92). Equivalently, a shift of the same size toward intermediation is associated with a 12.38 log percentage

point decline.

To summarize traffic responses among pure advertising and pure intermediation websites, I use empirical Bayes estimates that combine each website’s observed-minus-synthetic traffic change with its prediction from the composition model. This allows individual websites to depart from the fitted relationship while giving greater weight to more precise website-level estimates. I estimate an increase of 16.9 log percentage points (s.e. 7.8) for pure advertising websites and a decline of 11.1 log percentage points (s.e. 8.2) for pure intermediation websites. Here, pure refers to advertising’s share of in-scope revenue being one or zero, respectively.

The results of this study emphasize the ability of digital service taxes to distort market incentives in a digital context. They also provide evidence that turnover taxes on digital production can have unexpected and heterogeneous impacts on the value that digital platforms provide end users. In particular, the use of web traffic as an outcome broadens the scope of analysis to include effects beyond price incidence.

This paper is in conversation with a nascent empirical literature on the impacts of direct digital platform taxation. It’s possible to see the value of the traffic approach by taking stock of my estimates in the context of two recent papers that have focused on estimating price responses.

Combining these traffic estimates with existing evidence on advertising prices illustrates how the approach can shed light on platform inputs that are otherwise difficult to observe. [Acun et al. \(2026\)](#) estimate a 3.4 percent increase in advertisers’ effective costs per click following digital tax surcharges. Combining this price response with the decline in traffic to advertising-exposed nonliable websites yields an illustrative advertising-demand elasticity of approximately -2.4 . Together with the empirical Bayes estimated traffic increase of 3.91 log percentage points for Google, the calculation implies a decline in advertising intensity of approximately 12.15 log percentage points. Within the model, reduced advertising contributes approximately +12.15 log percentage points to traffic, offset by a content-and-services contribution of -8.24 log percentage points. These calculations impose strong model and cross-study assumptions, described in Section 5.

A similar calculation connects the intermediation results to evidence on consumer prices. [Langenmayr and Muddasani \(2025\)](#) estimate a French consumer log-price increase of 0.022, or 2.2 log percentage points, following the DST. Combining this price response with the empirical Bayes estimated 4.50 log percentage point traffic decline for Amazon gives a traffic-price elasticity of approximately -2.05 if the entire decline is attributed to prices. Within the model, this ratio provides an upper bound of 2.05 on the absolute price elasticity when reduced seller availability also discourages visits. The calculation does not separately identify the availability contribution.

This study also speaks to a closely connected research agenda consists of empirical research on indirect taxation of online commerce and platform-mediated transactions. [Agrawal and Fox \(2017, 2021\)](#) examine the collection and geographic assignment of consumption taxes in the digital economy, while studies of Airbnb and ride-hailing document how transaction taxes affect prices and market activity. [Bibler et al. \(2021\)](#) find that Airbnb tax enforcement raises prices paid by renters, and [Tarduno \(2025\)](#) finds substantial

pass-through of ridesharing taxes to passengers. These studies establish that taxation affects the terms on which users transact through digital platforms. This study both validates these results but also shows that the effects of digital taxation extend to a broader sample of liable websites.

A specific issue to digital taxation that has previously been understudied is that, as pointed out by (Hines, 2023), the multinational nature of digital service taxation, and in particular the public good nature of digital platform inputs, imply that any effects of digital taxation might have an impact on the provision of digital services in non-taxed jurisdictions. This will lead to inefficiently high digital tax rates. The analysis in this paper finds heterogeneous traffic responses in France, highlighting domestic behavioral effects alongside the government's ability to capture digital rents through the DST.

This paper relates to the broader literature on taxation in multi-sided markets, surveyed by Kind and Schjelderup (2025), and the literature on multi sided platforms as a whole (Rysman, 2009; Rochet and Tirole, 2003, 2006). Although I study platforms that serve multiple groups of users, an important intellectual contribution of the theoretical work is that none of the results rely on the distinctive pricing mechanisms emphasized in platform economics. In my framework, digital service provision is a cost of doing business, while the incidence of commissions and advertising costs reflects the balance of supply and demand in the market for the final good. The prediction that an advertising tax can increase traffic arises from the externality imposed on users who are shown advertisements they do not want to see. A tax that induces less advertising can reduce this externality and attract more visits, even as it lowers the return to providing services. These mechanisms guide the interpretation of the estimates, but the central empirical argument is broader: changes in traffic provide evidence that platforms change their operations in response to taxation.

Finally, this study has important implications for proposals to use advertising taxes to counteract the harmful effects of social media (Romer, 2019; Acemoglu and Johnson, 2024). The framework shows that reduced advertising can make a platform more attractive to users, potentially increasing participation. Normative analyses of targeted advertising taxes resembling the DST as a substitute for Pigouvian taxes on social media should therefore be careful to distinguish between its corrective effect on the advertising market vis a vis its potentially corrosive effect on over-consumption of digital services.

This paper proceeds as follows. In Section 2, I review the institutional context by providing a description of France's DST along with an exposition of liable platforms and their connection to nonliable websites. In Section 3, I outline a theoretical model that connects web traffic to unobservable platform inputs via user preferences. In Section 4, I explain my empirical approach, including a description of the data and the regressions that I would like to run and overview their results. In Section 5, I discuss the results in the context of the models in Section 3 and existing empirical research on DSTs. In Section 6, I conclude.

2 France's Digital Services Tax

France announced its intention to introduce a digital services tax in December 2018 amid ongoing international negotiations over the allocation of corporate taxing rights. The proposal reflected growing concerns

that existing international tax rules allowed multinational digital firms to generate substantial revenues from users in a country without reporting corresponding taxable profits there.

The French government's justification rested on the premise that user participation contributes to value creation by digital platforms and that existing international tax rules inadequately recognize this contribution. The DST was therefore intended to address perceived shortcomings in the taxation of multinational digital firms by taxing revenues associated with digital activity in France, pending a coordinated reform of international corporate taxation.

The distinction between the intended object of taxation—profits attributable to economic activity in France—and the statutory tax base—gross revenues allocated to France—is consequential. Digital platforms often rely on substantial investments in intangible assets, such as software, algorithms, and digital interfaces, that can serve users across multiple jurisdictions at relatively low incremental cost. Because the costs of developing these assets are shared across markets, attributing a portion of worldwide profits to any particular jurisdiction requires potentially contentious allocation rules. By contrast, advertising placements, transactions, and other revenue-generating activities can more readily be assigned to the locations of users.

The choice to tax revenues rather than profits has important implications for production incentives. A proportional tax on economic profits, with full deduction of economic costs, leaves a firm's optimal production decisions unchanged because it scales revenues and costs equally. A gross-revenue tax instead reduces the marginal return to production without reducing marginal costs, potentially discouraging otherwise profitable activity.

In a conventional model of ad valorem revenue taxation, zero marginal production costs imply that the tax leaves a firm's revenue-maximizing choices unchanged. This benchmark may appear particularly relevant to digital platforms, where the technological marginal cost of serving an additional user is commonly assumed to be negligible. However, the zero-marginal-cost premise, even where it holds, is applied to the wrong margin. Even if serving an additional unit of content or advertising costs almost nothing, producing the content that attracts users in the first place is far from free. This is the force of [Shapiro and Varian \(1999\)](#)'s observation that information goods are costly to produce but cheap to reproduce.

Moreover, sustaining user engagement requires ongoing investment in content, recommendations, and service quality rather than a single fixed expenditure. The economically relevant cost is therefore not simply the marginal cost of serving another visit, but the recurring cost of producing and maintaining the services that generate visits. A revenue tax can reduce the returns to these investments, even when the technological marginal cost of serving users is negligible. Whether platforms respond by reducing investment, altering service provision, or changing the allocation of activity across markets and competing websites is an empirical question that motivates this analysis.

2.1 Policy Details

The French DST imposes a 3% tax on gross revenues attributable to France from specified digital activities. For 2019, the tax applied to firms or corporate groups whose revenues from taxable services in 2018 exceeded both €750 million worldwide and €25 million in France. These thresholds are assessed using taxable-service revenues aggregated across entities within a corporate group, rather than total corporate revenues or profits.

The tax distinguishes two principal categories of digital services. Targeted advertising encompasses the placement of advertisements based on data collected from users of digital interfaces, as well as related advertising services and the sale of user data for advertising purposes. Digital intermediation encompasses services that enable users to interact or transact through a digital interface, with taxable revenues consisting of fees and other payments received in exchange for providing these services.

Because the DST is levied on gross revenues, firms cannot deduct the costs incurred in providing taxable services. The tax base may also include payments for complementary services that are not economically independent of access to or use of a digital interface, such as fulfillment services bundled with marketplace access.

The DST attributes revenues to France using a service-specific measure of French user activity. For each taxable service, the French tax base is calculated by multiplying worldwide taxable revenues by a French presence coefficient, which measures the share of relevant activity attributable to France. For targeted advertising, this coefficient equals the share of advertisements targeted toward users located in France, rather than the proportion of advertising expenditures attributable to France. For online marketplaces, it equals the share of transactions involving at least one user located in France. For other intermediation services, it is based on the share of active users whose accounts were opened from France.

Accordingly, a firm's DST liability can be expressed as

$$T = \tau \sum_{s \in \mathcal{S}} R_s \alpha_s, \quad (1)$$

where $\tau = 0.03$, R_s denotes worldwide taxable revenue from service s , and α_s is its French presence coefficient.

This equation underscores the underappreciated fact that the French digital service tax is not, in fact, strictly a tax on local revenues, but rather a tax on the French portion of global revenues, where the presence share is defined as a function of the firm's interactions with users rather than the location of revenues. In the context of the French government's justification for the tax, this may appear sensible, but it introduces additional distortions beyond those created by the decision to tax gross revenues rather than profits. To see this, consider the effective marginal tax on an advertisement sold in France.

$$t_F^{\text{eff}} = \tau \left[\alpha_A + (1 - \alpha_A) \frac{\bar{p}}{p_F} \right], \quad (2)$$

Here, α_A is the French share of worldwide advertising impressions, $\bar{p}$ is average worldwide revenue per impression, and p_F is the revenue from an additional impression shown in France. The effective marginal rate exceeds the statutory 3 percent when $p_F < \bar{p}$, falls below it when $p_F > \bar{p}$, and equals it when the two coincide. The difference arises because an additional French impression increases both worldwide advertising revenue and the share allocated to France. Although one impression changes the allocation share only slightly, that change applies to the entire worldwide revenue base.

Equation 1 also creates an implicit distortion on ads and intermediation that take place outside of France, this is because an additional foreign impression can reduce french tax liability when its revenue is below the worldwide average. Again, consider the effective marginal tax on an advertisement shown outside of France.

$$t_R^{\text{eff}} = \tau \alpha_A \left(1 - \frac{\bar{p}}{p_R} \right), \quad (3)$$

where all notation is as before except that p_R represents the return on an ad shown outside of France.

Estimating these effective marginal tax rates is difficult because neither service-specific presence ratios nor relative advertising returns are directly observed. Table 1 presents suggestive calibrations for Facebook, Instagram, and LinkedIn, combining Statista’s modeled French shares of worldwide users in 2018 with published CPM benchmarks. The implied marginal rates on advertisements shown in France range from 3.40 to 8.84 percent, compared with the statutory rate of 3 percent. Rates on advertisements shown in Germany and the United States are much smaller, with negative rates for Germany and generally positive rates for the United States, except for the Instagram comparison. These calculations are illustrative: user shares proxy impression shares, and the price benchmarks cover selected advertiser samples and different periods rather than worldwide advertising in 2018.

Table 1: Calibrated marginal French DST rates using Statista usage shares and CPM benchmarks

Product	CPM period	α_F (%)	q_F	Marginal rate (%)		
				France	Germany	US
Facebook	2015 Q4	1.47	0.336	8.84	-0.020	+0.011
Instagram	2015 Q4	1.75	0.881	3.40	-0.006	-0.017
Facebook	2016 Q2	1.47	0.499	5.96	-0.012	+0.006
Facebook, app installs	2016 Q2	1.47	0.482	6.17	-0.011	+0.009
LinkedIn, single image	2025–26	2.78	0.568	5.22	-0.021	+0.034

Notes: Rates are percentages of marginal advertising revenue; negative entries indicate implicit subsidies. α_F is the platform-specific ratio of modeled French to worldwide users in the 2018 Key Market Indicators estimates of [Statista \(2026a,b,c\)](#): Facebook 1.475%, Instagram 1.750%, and LinkedIn 2.783%. These audience shares proxy advertising-impression shares; they are not the parent-level traffic shares from the estimation sample. Statista’s geographic modeling also uses traffic information, and the activity-frequency, reference-date, and platform-scope definitions remain imperfectly documented. q_F is French CPM divided by the source’s multinational CPM benchmark. CPM denotes cost per thousand impressions. Sources for prices are [Salesforce Marketing Cloud \(2016a,b\)](#) for the 2015–16 benchmarks and [Kiin Labs \(2026\)](#) for the 2025–26 LinkedIn benchmarks. Calculations substitute source benchmarks for worldwide average prices and average prices for marginal prices. Periods in the table describe price evidence; all usage shares refer to 2018. These are conditional calibrations rather than measured 2018 tax rates; no sampling confidence intervals are available.

2.2 Timeline

The French DST emerged from earlier European efforts to reform the taxation of digital services. In March 2018, the European Commission proposed a common 3% tax on revenues from certain digital activities. However, the proposal failed to secure the unanimous support required for adoption. Following unsuccessful negotiations in December 2018, France announced that it would proceed unilaterally. On December 17, 2018, Finance Minister Bruno Le Maire announced that France would introduce a tax covering digital advertising, online marketplaces, and the commercialization of user data, with revenues taxable beginning January 1, 2019. The government introduced legislation in March 2019, and the DST was enacted in July 2019 with retroactive application to revenues earned from the beginning of the year.

Although the tax applied to 2019 revenues, collection and enforcement were subsequently affected by international negotiations. The first installment was payable in November 2019. In January 2020, France and the United States reached an agreement to defer certain subsequent payments while negotiations over international tax reform continued. These deferrals did not eliminate the underlying tax liability, and the deferred installments became payable in December 2020. The distinction between liability and collection is important because firms' incentives to adjust their behavior depend on expected tax obligations, not necessarily the dates on which payments are made.

The DST was also introduced as an interim measure pending a multilateral agreement on international corporate taxation. In October 2021, France joined an agreement establishing transitional arrangements for existing digital services taxes in anticipation of the implementation of the OECD's Pillar One framework. The agreement contemplated the eventual withdrawal of unilateral measures and conditional credits against future tax liabilities. These developments could affect firms' expectations about the duration of the DST, even while the tax remained in effect.

Figure 1 summarizes these developments, distinguishing the announcement and enactment of the tax, the timing of collections, and subsequent international agreements.

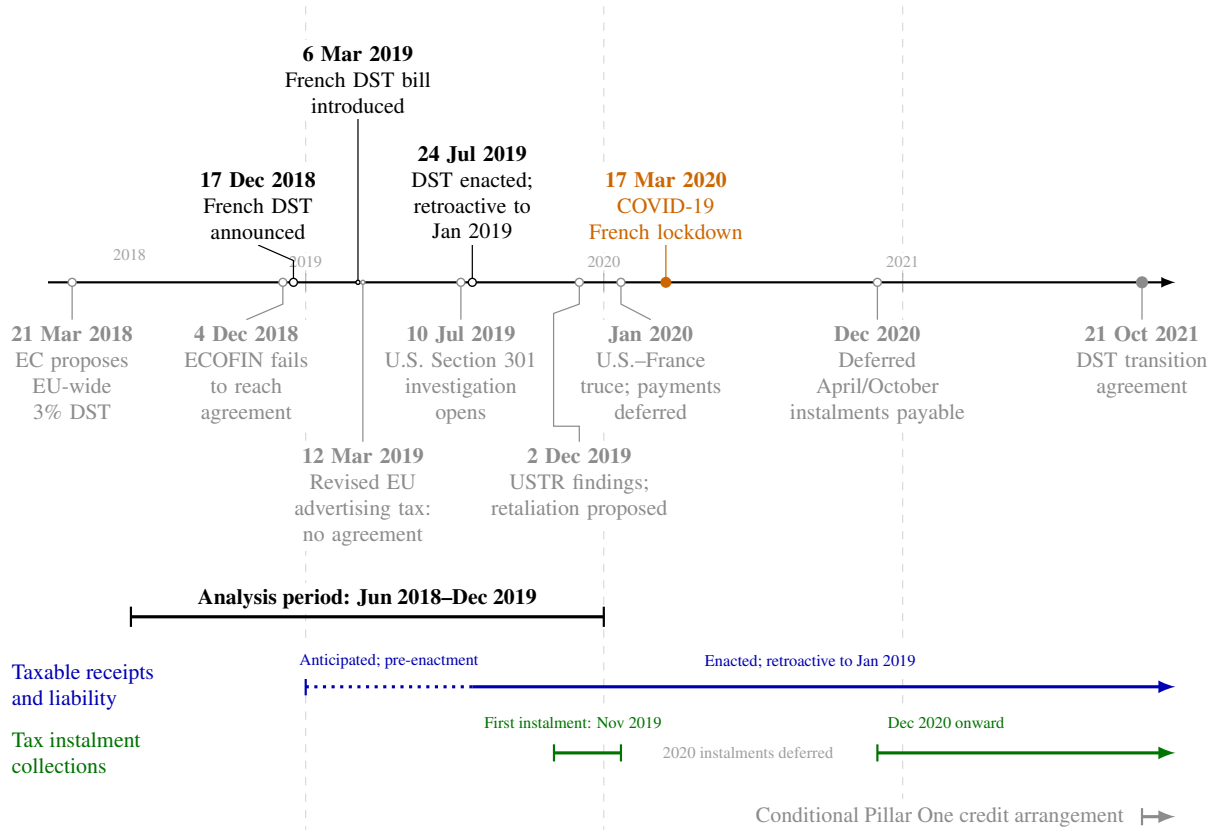


Figure 1: Timeline of the French digital services tax.

Notes: Event markers and period boundaries use proportional calendar-day spacing; labels are staggered for readability. Month-only payment events use the first of the month. The bracket covers June 1, 2018 through December 31, 2019 (right boundary: January 1, 2020). The dotted segment shows receipts subsequently covered by the law and anticipated liability before enactment; the solid segment shows enacted liability. The instalment-collection bars summarize instalment-collection periods, not continuous payments or all settlements. The first instalment was payable in November 2019. The optional deferral covered the April and October 2020 instalments, payable in December 2020; it did not cancel liability or imply that every other payment was suspended. The October 21, 2021 agreement provided conditional credits against future Pillar One Amount A liability; the specified interim credit period began January 1, 2022. Sources: European Commission (2018) and Council (2019); Assemblée nationale (2018); French Law No. 2019-759 (2019); DGFIP (2019; 2021); USTR (2021); and U.S. Treasury (2021).

2.3 Affected firms and their responses

The revenue thresholds used to determine corporate group-level liability make it so that only a handful of companies are thought to be subject to the French DST. In this section, I identify a group of 35 corporate groups that I believe are likely liable.

To start, I begin with a list of 27 corporate groups listed by the U.S. Trade Representative's report on France's DST (United States Trade Representative, 2019). Reasonable minds may disagree on the accuracy of this list, given the report's conclusion that the French DST was designed to unfairly target American corporations. Concerns in this regard can be separated into two forms of misclassification: false positives and false negatives.

To screen for false positives in the USTR classification, I collect fiscal-year 2018 consolidated rev-

enue data from publicly available corporate accounts and complement these with French-entity turnover records from Orbis.¹ Columns 1 and 2 of Table 2 report these measures. All 27 groups have total consolidated revenues above the global threshold. Among the 24 groups with positive recorded French-entity turnover, 17 exceed the local threshold.

Columns 3 and 4 improve on this rough test by showing revenues from advertising and intermediation, the two tax bases, taken from segment reports. This allows me to get a deeper view into the consolidated revenue shares by tax base. 23 of the 25 groups with available data pass the in-scope revenue thresholds. The two that do not are Match Group and Zalando. For Match group, the ad plus intermediation segment reports omit an unusual third category of subscription revenues that is likely in the liable tax base. Taking those into account raises its global revenues past the threshold. Zalando, on the other hand, is more likely to fail the global in-scope revenue tests, because its large size reflects the fact that most transactions that happen on their platform are first party.

The issue of false negatives in the USTR classification is more challenging because it requires assessing a large and not well defined set of candidate firms. In order to make this problem tractable, I create a candidate universe based on firms that share at least one in-scope industry code with one of the 27 firms in the USTR sample, then I select the subset of these firms that met the consolidated revenue thresholds and, when available, the French-specific revenue thresholds. I then review these candidates' business activities and available evidence on in-scope revenues, identifying eight additional corporate groups that are likely liable for the French DST.

In addition to the search for liable groups detailed above, I searched through the text of publicly available corporate filings and investor reports for mentions of the French DST. This search provided additional evidence of DST liabilities for four firms in this list: Booking, ANGI, TripAdvisor, and Match. See Appendix B for more details on this exercise.

Firms' announced responses show how the tax can change the terms offered to other platform participants. In August 2019, Amazon stated that certain seller fees for sales on Amazon France would increase by 3 percent from October 1, 2019 (Amazon, 2019). This was a proportional increase in fees, rather than a three-percentage-point increase in the commission rate. Apple's September 2020 notice stated that developer proceeds would be adjusted to account for France's 3 percent DST while posted App Store prices would remain unchanged, shifting the charge to developers through lower proceeds (Apple, 2020).

Advertising platforms also introduced explicit surcharges. Google added a 2 percent regulatory operating cost to advertisements served in France from May 1, 2021, describing it as covering part of the cost of complying with the French DST (Google, 2021). Meta later announced a 3 percent location fee for advertisements delivered to users in France, effective July 1, 2026 (Adegbola, 2026). These announcements document changes in contractual charges, while their ultimate incidence depends on subsequent responses by sellers, developers, advertisers, and users. Amazon's announced change falls within the paper's 2019 post-announcement window; the Apple, Google, and Meta changes illustrate subsequent adjustments.

¹Appendix A describes the data collection and its limitations.

The other thing to observe in Table 2 is the revenue share columns 5 through 7. This data is presented visually in Figure 2 for readability. Panel A of Figure 2 shows that although many firms operate almost exclusively in the in-scope markets, a substantial share of them have significant revenues in other businesses. Most often, this third line of business is direct retail sales of software, hardware, and other goods. There are also notable cases where the parent is a big player in the cloud computing business, and one case where the parent company’s revenues primarily reflect telecom income. Panel B shows how the firms map onto the two taxable revenue categories as a share of each firm’s DST-taxable income. I observe that about one-third of the firms only participate in advertising, one-third only participate in intermediation, and the final third fall somewhere in between.

The next task is to map corporate groups to websites owned by them before treatment. Table 3 lists these corporate groups and the websites they owned before treatment. It also reports November 2018 French traffic ranks and identifies additional websites with limited pre-treatment coverage.²

Table 4 reports the treated websites’ share of pre-treatment French and global traffic within the top 100, 1,000, and 2,000 website families. This information shows that despite the small number of corporate groups and websites considered liable for the French DST, they constitute a very large share of traffic to the top websites in France and the world. Finally, Panel B and Figure 3 summarize the relative “Frenchness” of websites in my sample, ranking them by their estimated French share of global traffic. These measures capture differences in geographic orientation within the sample rather than the French presence ratios used to calculate tax liability.

Table 3: Corporate groups and their sampled websites

Corporate group	Websites in the analysis sample	Other WTM websites with limited pre-treatment data
Airbnb	Airbnb (102)	–
Alibaba	AliExpress (11), Alibaba (350), Taobao (768)	Tmall (1,296), Youku (1,550), 1688 (1,745)
Alphabet	Google (1), YouTube (2)	Kaggle (743)
Amadeus	–	–
Amazon	Amazon (4), Twitch (37), IMDb (180), Gamepedia (212), Goodreads (441), CurseForge (470), Futhead (564), DPRReview (718), HearthPwn (918)	Souq (1,282), Mechanical Turk (2,197)
ANGI	–	–
Apple	–	–
Axel Springer	La Centrale (79), SeLogger (84), Logic-Immo (264), Welt (990), Computer Bild (1,147), Bild (1,174)	Caradisiac (147), Idealo (1,039), Finanzen (1,164), Business Insider (1,213), Immoweb (1,382), Immowelt (1,392), Auto Bild (1,522), meinestadt (1,604), Transfermarkt (1,688), StepStone (1,704), kaufDA (1,951), Immonet (1,960), Travelbook (2,212)
Booking Holdings	Booking.com (101), Rentalcars.com (808)	Agoda (870), Kayak (1,535), Momondo (2,212), Swoodoo (n.a.)
Cheil	–	–
ContextLogic	Wish (218)	–

Continued on next page

²Appendix A documents the evidence on likely liability and the links from firms to observed websites.

Table 3: Corporate groups and their sampled websites (continued)

Corporate group	Websites in the analysis sample	Other WTM websites with limited pre-treatment data
Criteo	–	–
Dentsu	–	–
eBay	eBay (8), mobile.de (379), eBay Kleinanzeigen (432), 2ememain (505), Marktplaats (800), Kijiji (906)	Gumtree (925), 2dehands (1,249), DBA (1,627), Motor-Talk (1,695)
Expedia	Hotels.com (515)	Expedia (1,327), Vrbo (2,034), FeWo-direkt (2,091)
Facebook	Facebook (3), Instagram (53)	–
Groupon	–	Groupon (1,439)
Interpublic	–	–
Just Eat	–	–
Match Group	Meetic (281), OkCupid (507), Plenty of Fish (538), Twoo (776)	Tinder (806), Match.com (2,171)

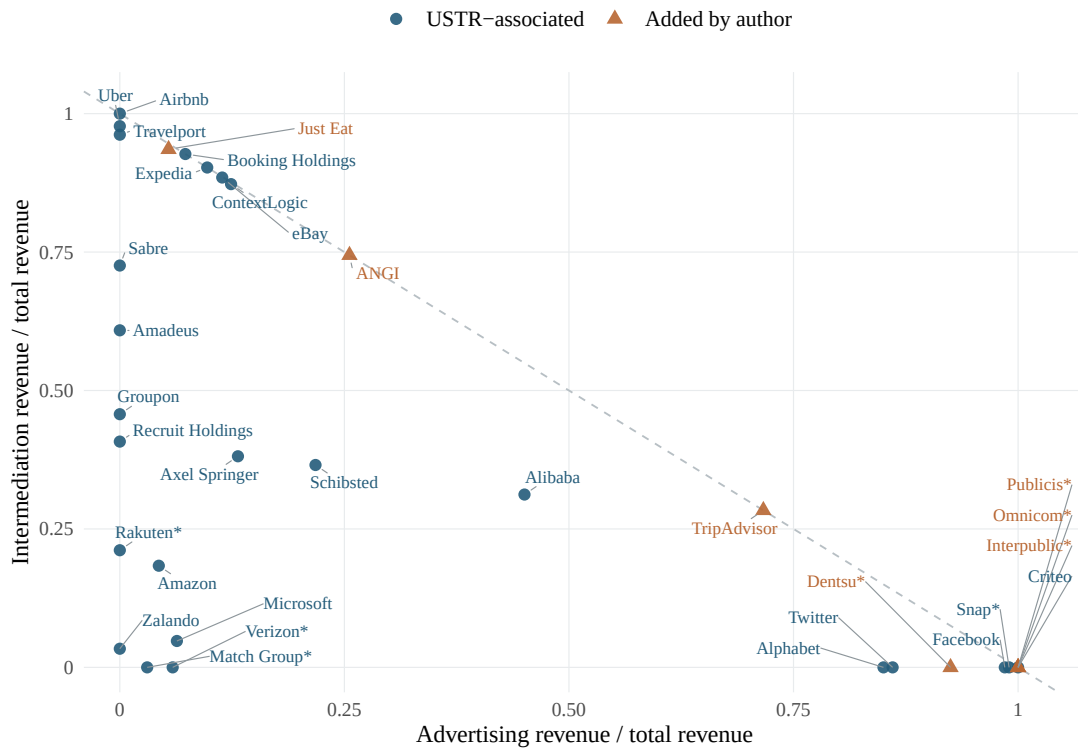
Continued on next page

Table 3: Corporate groups and their sampled websites (continued)

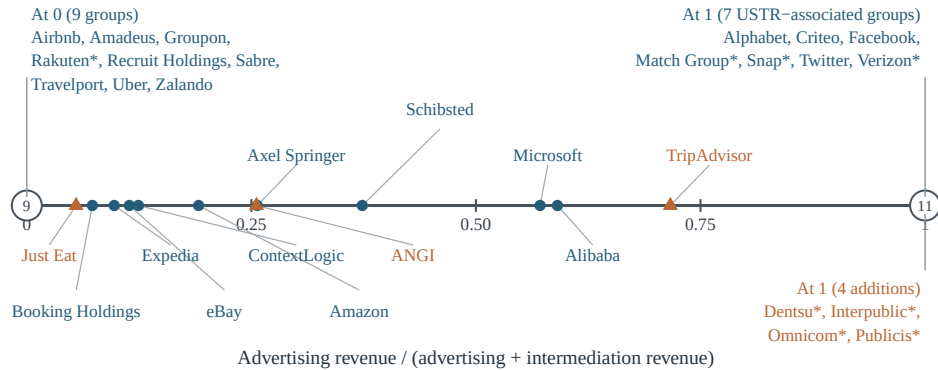
Corporate group	Websites in the analysis sample	Other WTM websites with limited pre-treatment data
Microsoft	Live (18), MSN (24), LinkedIn (34), Bing (91), SlideShare (529)	–
Omnicom	–	–
Publicis	–	–
Rakuten	Viki (548)	Rakuten (60), Infoseek (1,622)
Randstad	–	–
Recruit Holdings	Indeed (78)	Glassdoor (1,566)
Sabre	–	–
Schibsted	Leboncoin (9), Subito (766), Milanuncios (1,133)	Willhaben (1,306), InfoJobs (1,454), FINN (1,472), Blocket (1,583), Aftonbladet (1,606), OLX (1,756), VG (1,779), Shpock (1,824), Kufar (1,934), Tori (2,171), Jofogas (n.a.)
Snap	–	–
Travelport	–	–
TripAdvisor	TripAdvisor (157)	–
Twitter	Twitter (15)	–
Uber	Uber (665)	–
Verizon	Yahoo (13), Huffington Post (138), Tumblr (143), AOL (318)	TechCrunch (1,065), Engadget (1,135)
Zalando	Zalando (95)	–

Notes: The analysis sample contains 55 website families under 20 of the 35 corporate groups. Ownership follows the pre-treatment mapping described in Appendix A. Parentheses give November 2018 French traffic ranks among all 2,230 website families observed in WhoTracksMe that month (1 = highest traffic), after summing observed member-domain popularity. Families combine country-specific domains; ranks are not restricted to liable websites. The last column lists 49 additional families that pass the recorded activity review but have only one to three positive French observations in June–November 2018 and are outside the selected 55-family roster. It excludes websites rejected for other reasons and those first observed after treatment. “n.a.” means no November observation; a dash means no website in that column. Sources: WhoTracksMe (Karaj et al., 2018) and United States Trade Representative (2019); see Appendix A.

A. Advertising and intermediation shares of total revenue



B. Advertising share of advertising plus intermediation revenue



FY2018. Equal-size corporate-group markers; 32 groups with available shares.

Apple, Randstad and Cheil are omitted. * indicates a proxy or classification limitation; see figure notes.

Figure 2: Corporate revenue composition, FY2018

Notes: Panel A plots FY2018 advertising and intermediation revenues as shares of FY2018 consolidated revenue; the dashed line indicates that the two sum to total revenue. Panel B places advertising's share of their combined revenue on a number line; endpoint circles report group counts. Labels are offset for readability. Blue circles denote USTR-associated groups; orange triangles denote the author's additions. The figure preserves the classifications in Table 2; Appendix A.5 documents the company filings, Orbis data (Bureau van Dijk, a Moody's Analytics company, 2025) and revenue classifications. Asterisks flag limitations: Match's dating subscriptions are coded outside intermediation; Rakuten's advertising is not separately allocated; Verizon uses Oath revenue; Snap uses its rounded 99% advertising share; agency revenues include traditional advertising. Apple, Randstad, and Cheil are omitted because their splits are missing.

Table 2: Corporate revenues and business composition, FY2018

Corporate group	Revenue (EUR millions)				Revenue shares (%)			
	Consolidated total	French entities	Ads	Inter-mediation	Ads/total	Interm./total	Ads/(ads+interm.)	French/total
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. USTR-associated groups (27)</i>								
Airbnb	3,099.6	14.3	0.0	3,099.6	0.0	100.0	0.0	0.46
Alibaba	48,385.1	5.4	21,796.6	15,101.6	45.0	31.2	59.1	0.01
Alphabet	116,125.7	412.2	98,725.4	0.0	85.0	0.0	100.0	0.35
Amadeus	4,935.7	1,713.8	0.0	3,004.3	0.0	60.9	0.0	34.72
Amazon	197,663.8	735.1	8,579.2	36,280.0	4.3	18.4	19.1	0.37
Apple	225,424.9	930.8	—	—	—	—	—	0.41
Axel Springer	3,180.7	110.3	418.3	1,212.5	13.2	38.1	25.6	3.47
Booking Holdings	12,329.9	35.7	899.7	11,430.2	7.3	92.7	7.3	0.29
ContextLogic	1,466.6	—	181.6	1,279.9	12.4	87.3	12.4	—
Criteo	1,952.4	875.7	1,952.4	0.0	100.0	0.0	100.0	44.85
eBay	9,120.7	6.6	1,039.7	8,068.3	11.4	88.5	11.4	0.07
Expedia	9,525.6	158.2	926.8	8,598.7	9.7	90.3	9.7	1.66
Facebook	47,392.7	390.4	46,692.5	0.0	98.5	0.0	100.0	0.82
Groupon	2,237.9	41.9	0.0	1,023.2	0.0	45.7	0.0	1.87
Match Group	1,468.2	27.9	44.7	0.0	3.0	0.0	100.0	1.90
Microsoft	93,668.5	1,439.7	5,951.5	4,463.6	6.4	4.8	57.1	1.54
Rakuten	8,474.1	—	0.0 ^c	1,792.6	0.0	21.2	0.0	—
Randstad	23,812.0	3,938.8	—	—	—	—	—	16.54
Recruit Holdings	17,777.4	741.5	0.0	7,248.7	0.0	40.8	0.0	4.17
Sabre	3,282.1	8.3	0.0	2,381.8	0.0	72.6	0.0	0.25
Schibsted	1,877.8	417.7	409.1	686.3	21.8	36.5	37.3	22.24
Snap	1,001.9	46.7	≈991.9	0.0	≈99.0	0.0	100.0	4.66
Travelport	2,165.2	11.4	0.0	2,082.9	0.0	96.2	0.0	0.53
Twitter	2,582.2	13.3	2,221.5	0.0	86.0	0.0	100.0	0.52
Uber	9,565.5	58.2	0.0	9,349.0	0.0	97.7	0.0	0.61
Verizon	111,070.5	4.7	6,535.4 ^a	0.0	5.9	0.0	100.0	< 0.01
Zalando	5,387.9	—	0.0	181.4	0.0	3.4	0.0	—
<i>Panel B. Groups added by the author (8)</i>								
ANGI	961.0	—	245.7	715.3	25.6	74.4	25.6	—
Cheil	2,716.6 ^b	54.7	—	—	—	—	—	2.01
Dentsu	7,835.8	312.6	7,246.6 ^a	0.0	92.5	0.0	100.0	3.99
Interpublic	8,245.1	191.5	8,245.1 ^a	0.0	100.0	0.0	100.0	2.32
Just Eat	879.4	43.0	47.7	822.7	5.4	93.6	5.5	4.89
Omnicom	12,977.6	493.3	12,977.6 ^a	0.0	100.0	0.0	100.0	3.80
Publicis	9,951.0	1,155.6	9,951.0 ^a	0.0	100.0	0.0	100.0	11.61
TripAdvisor	1,370.7	—	982.0	388.7	71.6	28.4	71.6	—

Notes: Columns (1), (3), and (4) are FY2018 consolidated revenues from company filings, in euros. Column (2) is turnover of the group's French entities from unconsolidated Orbis accounts; it is not sales to French users or the DST base. Columns (5)–(7) are $100 \times (3)/(1)$, $100 \times (4)/(1)$, and $100 \times (3)/[(3) + (4)]$; column (8) is $100 \times (2)/(1)$, calculated before rounding. Dashes indicate unavailable splits or accounts; zeros indicate no revenue allocated to the category. ^aIncludes traditional advertising (agencies) or uses Oath revenue (Verizon). ^bFrom Orbis. ^cNot separately allocated. $\approx$ Approximated from a reported share. Sources: company filings, Orbis (Bureau van Dijk, a Moody's Analytics company, 2025), and United States Trade Representative (2019). Appendix A.5 documents the sources, revenue classification, exchange rates, and each adjusted entry.

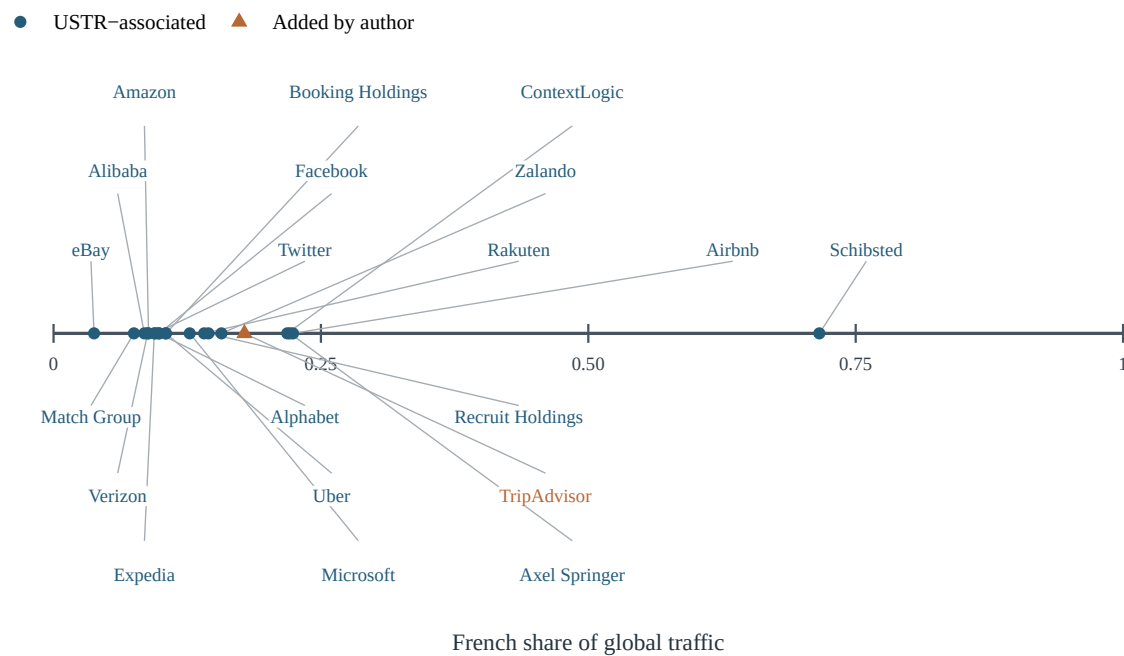


Figure 3: French shares of global traffic in WTM data

Notes: Each marker places one of the 20 treated corporate groups on a common number line from 0 to 1. Traffic covers the group's selected analysis-sample website families in September–November 2018. French popularity is calibrated monthly assuming that 90% of Amazon.fr's global traffic originates in France; each group share divides summed calibrated French popularity by summed global popularity across its sampled families and months. Shares describe sampled websites rather than all group properties. Blue circles identify USTR-associated groups; the orange triangle identifies the group added by the author. Labels are offset solely for readability. Missing observations are omitted. Source: author's calculations on WhoTracksMe data (Karaj et al., 2018).

Table 4: Pre-treatment traffic coverage and French shares of global traffic in the WTM data

Panel A. Treated traffic within the top N website families

Top N families	France		Global	
	Treated families	Traffic share (%)	Treated families	Traffic share (%)
100	19	56.49	24	70.25
1,000	52	44.15	46	51.90
2,000	55	43.78	52	47.52

Panel B. Estimated French share of global traffic (%)

Sample	n	Mean	SD	Min.	P25	Median	P75	Max.
Website families	55	16.28	26.00	0.23	4.50	8.81	14.28	96.77
Corporate groups	20	15.50	14.18	3.79	9.29	10.50	16.24	71.60

Notes: Panel A uses November 2018 WhoTracksMe popularity and the family definitions in Table 3. France (2,230 families) and the source’s global series (2,382 families) are ranked independently; both use the same 55 treated families. Panel A divides treated popularity within each top N by all popularity within that top N ; cutoff ties break alphabetically. Panel B uses September–November 2018. In month t , French popularity is multiplied by $k_t = 0.9 G_{\text{amazon.fr},t} / F_{\text{amazon.fr},t}$, assuming 90% of Amazon.fr’s global traffic originates in France. Each share is $100 \sum_t k_t F_{it} / \sum_t G_{it}$. Group shares first sum traffic across the group’s sampled families; they do not cover all group properties. Rows weight families or groups equally. SD is the sample standard deviation (percentage points in Panel B); quartiles use linear interpolation. Missing observations are omitted. Source: WhoTracksMe (Karaj et al., 2018) and the treatment roster documented in Appendix A.

2.4 Advertising Connections beyond Liable Platforms

The reach of liable platforms extends beyond the websites they own. Nonliable websites may purchase advertising from these platforms to attract visitors, supply advertising inventory through them, or do both. These connections provide channels through which the DST can affect websites outside its statutory scope.

Although historical advertising campaign records are unavailable, embedded third-party technologies reveal potential connections between websites and advertising providers. On an advertiser’s website, JavaScript tags and tracking pixels can record visits and actions such as viewing a product or completing a purchase. Cookies and identifiers transmitted with these requests can allow advertising providers to associate these actions with earlier ad interactions or use them to select subsequent advertisements. For example, Facebook Pixel supports retargeting based on a visitor’s interactions with an advertiser’s website, while Bing Ads’ Universal Event Tracking supports conversion measurement and remarketing. My database contains the broader Facebook and Bing Ads tracker identities, as well as Google AdServices and Criteo.³

Other technologies support the sale and delivery of advertising on third-party websites. The database identifies Google Syndication, Amazon Advertising, AppNexus, and Outbrain, among others. These services include publisher advertising networks, bidding infrastructure, and sponsored recommendations.

³Facebook’s April 2018 description (Facebook, 2018) gives a retail retargeting example using Facebook Pixel. Microsoft’s UET documentation (Microsoft, 2026) describes conversion tracking and remarketing; the latter is current product documentation, rather than a reconstruction of a particular website’s November 2018 implementation.

Some providers serve both advertisers and publishers, so a provider’s presence alone does not determine whether the host website buys advertising, sells advertising, or uses a related service.⁴

Not every third-party connection establishes advertising participation. A request for an image or another externally hosted resource can disclose a visitor’s presence and, depending on the request, transmit an identifier. But general content-delivery services, standalone analytics, and tag-management containers also support activities unrelated to advertising. For the empirical measures, I use WTM’s *advertising* category to identify advertising-related technologies, excluding other categories and the former custom narrow/broad overrides.

These technologies can support different sides of the advertising market: displaying or monetizing ads on the host website, or tracking conversions and enabling remarketing associated with ads shown elsewhere. The provider-level identities recorded in the data generally do not distinguish these uses. A connection to an advertising network may also reflect measurement or identifier matching, so its presence does not by itself establish that the host displays ads. Likewise, conversion-tracking infrastructure does not establish that the host purchased advertising. I therefore interpret these observations as advertising-related technical connections without assigning each host to an advertiser or publisher role.

Table 5 reports November 2018 coverage among the paper’s nonliable website families. Panel A lists the fifteen leading advertising-related technologies and their corporate owners, ranked by the share of global nonliable traffic accounted for by families with a published detection. The columns show the percentage of nonliable families with a detection and their shares of global and French-origin nonliable traffic. Panel B reports combined coverage across all advertising-related technologies, separating liable-owned from nonliable-owned providers and unresolved ownership. A detection on any member domain identifies a connection for its family.

⁴Amazon’s January 2018 announcement ([Amazon, 2018](#)) describes its publisher bidding service and its availability in France. Criteo’s 2018 annual report ([Criteo, 2019](#), pp. 15–16) describes its advertiser and publisher services.

Table 5: Advertising-related technologies among nonliable websites, November 2018

Technology	Corporate owner	Sites (%)		Traffic to identified sites (%)	
		Global	France	Global	France
<i>A. Top fifteen advertising-related technologies</i>					
Google	Alphabet*	84.0	52.4	88.9	83.9
DoubleClick	Alphabet*	73.8	43.7	80.1	79.3
Facebook	Facebook*	59.7	42.8	62.7	64.6
Google AdServices	Alphabet*	31.0	18.7	37.8	41.4
Google Syndication	Alphabet*	24.6	9.7	32.3	32.0
AppNexus	AT&T	24.3	12.7	29.7	43.2
Yandex	Yandex	14.1	6.3	24.2	4.7
Criteo	Criteo*	19.3	10.7	23.5	30.8
Amazon Advertising	Amazon*	15.7	7.7	22.5	21.0
Quantcast	Quantcast	12.4	8.0	18.7	7.6
Adobe Audience Manager	Adobe	14.4	8.0	18.5	20.3
Bing Ads	Microsoft*	17.2	9.6	16.6	22.7
TradeDesk	The Trade Desk	10.7	2.5	14.7	5.9
MediaMath	MediaMath	8.9	2.7	13.6	7.5
Moat	Oracle	4.2	1.4	13.2	1.0
<i>B. Combined provider coverage</i>					
Any advertising-related technology		94.8	77.7	96.9	93.5
Liabile-owned providers		88.9	67.9	92.9	90.6
Nonliable-owned providers		76.2	52.2	79.8	70.9
Unresolved ownership		15.2	4.7	32.4	26.1

Notes: All values are percentages. Both panels restrict technologies to WTM’s advertising category. The universe are all 638 nonliable website families observed in November 2018 either in the French traffic extract or the global traffic extract. A published detection on any member domain identifies a website-level connection. Panel A is ranked by global nonliable traffic coverage. Panel B uses all advertising-related technologies, including those outside the top fifteen, counting a family once per row. Rows overlap and must not be added. Corporate owners are as of January 2019; * denotes an owner classified as likely liable for the French DST in this study. Source: WhoTracksMe site and site-tracker panels (Karaj et al., 2018); saved April 2026 WTM category metadata for advertising categorization.

3 Digital Tax Incentives and Website Popularity

In this section, I connect the economic incentives created by the digital services tax to observed website traffic. The central idea is that the tax can change platform choices that are unobserved in the data but affect users' willingness to visit. I first describe this connection without specifying how platforms produce or monetize their services. I then introduce two parsimonious models as frameworks for interpreting the traffic response. It's worth emphasizing that the interpretation I assign to the causal estimates does not depend on the specifics of either model. These are instead meant to clarify the economic mechanisms that could produce them.

Let $\mathbf{q}_i$ denote a vector of characteristics of website i that result from platform choices. They may include the content available, the convenience of accessing it, advertising intensity, and other amenities or disamenities. Users' choices about how often to visit therefore reveal their demand for the bundle of characteristics the platform provides.

$$\mathcal{L}_i(\mathbf{q}_i) = \arg \max_{L_i \geq 0} v_i(L_i; \mathbf{q}_i). \quad (4)$$

Here, v_i denotes the user's utility from visiting website i , given its characteristics. This formulation allows a site-specific satiation point without introducing an explicit consumption cost.

When a new digital service tax τ distorts the platform's choices, the user's demand for page loads changes according to:

$$\frac{\partial \mathcal{L}_i}{\partial \tau} = \sum_k \frac{\partial \mathcal{L}_i}{\partial q_{ik}} \frac{dq_{ik}}{d\tau}. \quad (5)$$

The response combines two elements: how the platform changes each characteristic and how users' demand responds to that characteristic. This relationship does not require a particular model of platform production or monetization.

Under these assumptions, a nonzero causal effect on traffic implies that the tax induced a change in at least one platform choice that matters for user demand. A zero effect admits several interpretations. The platform may leave its choices unchanged; users' visits may respond little to the characteristics that change; or the traffic effects of different adjustments may offset one another. Traffic alone therefore cannot identify which choices changed or establish that the platform did not respond.

The next two subsections illustrate this argument using models of an advertising-funded social media platform and an e-commerce platform that intermediates third-party transactions. These models connect the two statutory tax bases to possible platform responses. They provide interpretations of the estimated traffic effects rather than claims about the particular adjustments platforms made.

In what follows, note that I assume that $\bar{p} = p_F$ in Equation 2, so that the effective marginal rate is equal to the statutory marginal rate of 3%. This is without loss of generality, as I may consider τ in what follows as the effective marginal rate.

3.1 Social Media Platforms

Now I turn to the promised hypothetical social media platform that invests in engaging content to attract users and monetizes the resulting attention by showing them advertisements. The platform incurs two types of expenditure: g_U funds services that attract users, while g_A funds the sale and delivery of advertising. A tax on advertising revenue reduces the return to both. Its effect on traffic depends on how the platform adjusts the services it provides and the advertising users encounter.

Let L denote page loads and A advertising impressions. Advertising intensity, $a = A/L$, measures impressions per page load. Traffic satisfies

$$L = \mathcal{L}(a, g_U), \quad \mathcal{L}_a < 0, \quad \mathcal{L}_{g_U} > 0, \quad \mathcal{L}_{g_U g_U} \leq 0.$$

User-facing expenditure attracts traffic with diminishing returns. Advertising discourages visits because it interrupts or competes with the content users seek. This assumption describes a content-consumption setting; advertisements that help users find products can have a different effect.

Delivering impressions requires advertiser-facing expenditure according to

$$A = F(g_A), \quad F'(g_A) > 0, \quad F''(g_A) \leq 0.$$

The platform faces inverse advertising demand $p_A = P_A(A)$, with $P'_A(A) < 0$. Advertisers pay less per impression as the platform supplies more impressions. I hold impression quality and audience composition fixed, so their willingness to pay depends on the number of impressions rather than separately on advertising intensity or user-facing expenditure. Appendix C.1 derives this demand curve from a downstream product market.

Let $\tau \in [0, 1)$ denote a tax on advertising revenue. The platform chooses advertising intensity and both expenditures to maximize

$$\begin{aligned} \max_{a, g_U, g_A \geq 0} \quad & (1 - \tau)P_A(A)A - g_A - g_U \\ \text{subject to} \quad & A = F(g_A) = a\mathcal{L}(a, g_U). \end{aligned}$$

This characterization of advertising production serves as a realistic portrait of my toy social media platform. Note that the characterization of the firm's costs is meant to emphasize the idea that advertising and traffic are best thought of as downstream outcomes of ex-ante investment decisions made by digital platforms. Doing so helps to avoid confusion about the marginal cost structure of digital production.

Although for pedagogical reasons it is useful to characterize traffic and advertising as an outcome of unobservable platform choices, in practice it is mathematically convenient to equivalently express the optimization problem simply as a function of A . To do so, start by noting that for a given number of impressions, revenue is fixed. The platform therefore chooses the least costly combination of advertising intensity and

user-facing expenditure that supports those impressions. Define this user-facing expenditure as

$$U(A) = \min_{a, g_U \geq 0} \{g_U : a\mathcal{L}(a, g_U) = A\}.$$

Advertiser-facing expenditure is $F^{-1}(A)$, so total required expenditure is $C_A(A) = F^{-1}(A) + U(A)$. This characterization of the cost function is especially useful conceptually, as it emphasizes that users are a costly input into the production of advertising.

At the expenditure-minimizing allocation, the first-order and envelope conditions derived in Appendix C.2 give

$$\mathcal{L} + a\mathcal{L}_a = 0, \quad U'(A) = \frac{1}{a\mathcal{L}_{g_U}} > 0.$$

The first condition balances the extra impressions from showing another advertisement on each page against the impressions lost when users visit less. The second states that supporting more impressions requires more expenditure on the services that attract users.

The first expression is particularly important, because it implies a unit elasticity of traffic with respect to advertising intensity. This fact connects to the idea that a zero effect could imply that users are indifferent to advertising. This condition shows that expenditure minimization requires that users be in the unit elastic region of demand with respect to ads.

The platform's problem of how many impressions to deliver can then be written as

$$\max_{A \geq 0} \Pi(A; \tau), \quad \Pi(A; \tau) = (1 - \tau)P_A(A)A - C_A(A),$$

over feasible volumes with nonnegative advertising prices. Its first-order condition is

$$(1 - \tau) [P_A(A) + AP'_A(A)] = C'_A(A).$$

This expression states the optimality condition that after-tax marginal advertising revenue equals the marginal expenditure needed to supply another impression. A higher tax lowers this return and reduces the chosen volume. Here I obtain the first comparative static expression for total advertising:

$$\frac{dA}{d\tau} = \frac{P_A(A) + AP'_A(A)}{\Pi_{AA}(A; \tau)} < 0.$$

The numerator is positive because marginal expenditure is positive; the denominator is negative by the strict second-order condition. Fewer impressions require less advertiser-facing expenditure and less user-facing expenditure; Appendix C.2 derives these tax responses. In particular,

$$\frac{dg_U}{d\tau} = \frac{1}{a\mathcal{L}_{g_U}} \frac{dA}{d\tau} < 0.$$

Because of the decline in both ads and user-facing investments, the decline in impressions need not mean fewer advertisements per page. Appendix C.3 derives the response of advertising intensity to the reduction in user-facing expenditure:

$$\frac{da}{d\tau} = -\frac{\mathcal{L}_{gU} + a\mathcal{L}_{agU}}{2\mathcal{L}_a + a\mathcal{L}_{aa}} \frac{dg_U}{d\tau}.$$

The denominator is negative at the conditional expenditure minimum. Intensity therefore falls when $\mathcal{L}_{gU} + a\mathcal{L}_{agU} > 0$: greater expenditure on user services raises the marginal impression gain from increasing advertising intensity, so reducing that expenditure makes lower intensity optimal. If this expression is negative, intensity rises.

Traffic combines the effects of these two adjustments:

$$\frac{dL}{d\tau} = \underbrace{\mathcal{L}_a \frac{da}{d\tau}}_{\text{advertising intensity}} + \underbrace{\mathcal{L}_{gU} \frac{dg_U}{d\tau}}_{\text{user services}}. \quad (6)$$

Reduced expenditure on user services discourages visits. When advertising intensity also falls, fewer advertisements make the platform more attractive and offset part of that loss. Traffic rises only if this benefit exceeds the loss from reduced services. Equivalently, because $L = A/a$, intensity must fall proportionately more than total impressions. Appendix C.3 gives the corresponding condition on the derivatives of the traffic function.

The point here is that the predicted effect of the tax on traffic to a social media platform is ambiguously signed. In addition, under this framework, a zero effect on traffic is only consistent with offsetting effects of the tax on advertising and user content inputs.

3.2 Transaction Intermediation

In this section, I consider the effects of digital taxation on a hypothetical e-commerce intermediary. Relative to the social media case, this setting emphasizes two channels through which the tax can affect traffic: changes in the prices buyers pay and changes in the sellers available on the platform.

Let Q denote completed transactions, p_B the price paid by buyers, and p_S the amount sellers receive per transaction. Buyer demand and aggregate seller supply imply

$$p_B = D^{-1}(Q), \quad p_S = S^{-1}(Q),$$

with $(D^{-1})'(Q) < 0$ and $(S^{-1})'(Q) > 0$. The platform's fee is the difference between these prices:

$$p_Q = P_Q(Q) \equiv D^{-1}(Q) - S^{-1}(Q).$$

Higher fees reduce transaction volume: $P'_Q(Q) < 0$. The division of the fee between higher buyer prices and lower seller compensation depends on the slopes of demand and supply.

To distinguish availability from completed transactions, let

$$N = H(p_S), \quad H'(p_S) > 0,$$

denote the number of active sellers. Higher compensation makes participation worthwhile for more sellers, expanding the set of offers available to users. Aggregate supply $S(p_S)$ incorporates both seller participation and the quantities supplied by participating sellers. The participation response is an additional assumption; upward-sloping aggregate supply alone does not imply it.

Intermediation expenditure, g_Q , funds transaction processing, buyer and seller support, and other resources needed to facilitate purchases. Supporting transactions requires

$$Q = F(g_Q), \quad F'(g_Q) > 0, \quad F''(g_Q) \leq 0.$$

The required expenditure is therefore $C_Q(Q) = F^{-1}(Q)$, with $C'_Q(Q) > 0$. The model treats this expenditure as a cost of supporting transaction volume. Its effects on traffic operate through the buyer prices and seller participation associated with that volume.

Let $\tau \in [0, 1)$ denote a tax on fee revenue. The platform chooses transaction volume to maximize

$$\max_{Q \geq 0} \Pi(Q; \tau), \quad \Pi(Q; \tau) = (1 - \tau)P_Q(Q)Q - C_Q(Q), \quad (7)$$

over feasible volumes with nonnegative fees. The taxable revenue is the platform's fee income, $p_Q Q$. I consider smooth, positive interior solutions satisfying the strict second-order condition. Appendix D.1 derives the expenditure conditions and transaction-volume response.

The platform equates after-tax marginal fee revenue with the marginal expenditure required to support another transaction:

$$(1 - \tau) [P_Q(Q) + QP'_Q(Q)] = C'_Q(Q).$$

A higher tax lowers the return to additional transactions while leaving their expenditure requirements unchanged. The chosen volume falls:

$$\frac{dQ}{d\tau} = \frac{P_Q(Q) + QP'_Q(Q)}{\Pi_{QQ}(Q; \tau)} < 0.$$

The numerator is positive by the first-order condition, since marginal expenditure is positive. The denominator is negative by the strict second-order condition.

Lower transaction volume requires less intermediation expenditure, as derived in Appendix D.1, and corresponds to a higher fee. Appendix D.2 derives the fee and price responses. Along buyer demand and seller supply,

$$\frac{dp_B}{d\tau} = (D^{-1})'(Q) \frac{dQ}{d\tau} > 0, \quad \frac{dp_S}{d\tau} = (S^{-1})'(Q) \frac{dQ}{d\tau} < 0.$$

Buyers pay more and sellers receive less. Lower seller compensation also reduces participation:

$$\frac{dN}{d\tau} = H'(p_S) \frac{dp_S}{d\tau} < 0.$$

Thus, the tax reduces both completed transactions and the number of sellers available to users; Appendix D.2 derives the participation response under the separate assumption $H' > 0$.

Traffic includes browsing and comparison shopping as well as completed purchases. Let L denote page loads. Traffic depends on buyer prices and seller availability:

$$L = \tilde{\mathcal{L}}(p_B, N).$$

Assume that higher buyer prices discourage visits and greater seller availability attracts them:

$$\tilde{\mathcal{L}}_{p_B} < 0, \quad \tilde{\mathcal{L}}_N > 0.$$

These are assumptions about the net response of page loads, including visits that do not result in a purchase. In particular, the second assumption treats seller participation as an amenity for users, rather than equating availability with the number of completed transactions.

The traffic response is

$$\frac{dL}{d\tau} = \underbrace{\tilde{\mathcal{L}}_{p_B} \frac{dp_B}{d\tau}}_{\text{buyer prices}} + \underbrace{\tilde{\mathcal{L}}_N H'(p_S) \frac{dp_S}{d\tau}}_{\text{seller availability}} < 0.$$

Higher buyer prices discourage visits, while reduced seller participation makes the platform less attractive for browsing and purchasing. Both channels lower traffic under the maintained assumptions.

The negative traffic prediction depends on these responses. Appendix D.2 derives the more general condition: traffic falls when

$$\tilde{\mathcal{L}}_{p_B} (D^{-1})'(Q) + \tilde{\mathcal{L}}_N H'(p_S) (S^{-1})'(Q) > 0,$$

so that traffic increases with transaction volume along the market equilibrium. The sign assumptions above are sufficient for this condition. If higher prices or reduced availability instead lead users to search more intensively, page loads need not fall with transactions.

4 Estimating Traffic Responses

The core of this paper's empirical strategy consists of estimating the response to the French digital services tax by comparing changes in French-origin traffic between websites owned by liable firms and nonliable websites. The comparison asks whether traffic to liable websites changes after the December 2018 announcement relative to traffic to websites outside the tax's scope. I compare September–November 2018

with January–December 2019, excluding December 2018 because the announcement occurs during that month. Ending the comparison in 2019 avoids combining the initial response with changes in internet use during the COVID-19 pandemic.

My preferred specification compares liable website families with synthetic controls constructed from nonliable families without observed advertising exposure. I also report comparisons using advertising-exposed and pooled nonliable websites to assess how the choice of comparison group affects the estimates.

The baseline difference-in-differences specification is

$$y_{it} = \alpha_i + \lambda_t + \beta (L_i \times Post_t) + \varepsilon_{it}, \quad (8)$$

where i indexes website families and t indexes calendar months. The outcome y_{it} is the logarithm of family i 's French-origin traffic relative to Wikipedia in month t . The indicator L_i identifies families owned by liable firms, and $Post_t$ equals one in 2019. Website-family fixed effects, α_i , absorb persistent differences in traffic across families, while month fixed effects, λ_t , absorb monthly changes common to the sample. The coefficient β measures the change in log traffic to liable websites relative to nonliable websites.

I estimate the regression by ordinary least squares, giving each observed family-month equal weight. The estimate therefore describes differential traffic growth across website families rather than the change in their combined traffic volume. I calculate standard errors by resampling whole parent firms for liable websites and whole website families for nonliable websites, retaining all websites and months belonging to each sampled block. This allows dependence across liable websites under the same parent and over time within each block. The full-pipeline procedure below reestimates the synthetic weights in every draw.

I examine potential spillover effects within the nonliable group using the presence of advertising-related tracking technologies. I compare traffic changes between nonliable website families with and without observed advertising exposure, with advertising exposure defined in Subsection 4.3:

$$y_{it} = \alpha_i + \lambda_t + \delta (E_i \times Post_t) + \varepsilon_{it}, \quad (9)$$

where the sample contains only nonliable website families and E_i indicates an observed connection to a liable-owned WTM advertising technology in the French panel during September–November 2018. The estimation period and weighting follow the baseline specification; inference resamples full family histories separately within the exposed and residual groups.

The coefficient δ measures the post-announcement change in log traffic among nonliable families with observed exposure relative to those without observed exposure. A causal spillover interpretation requires that these groups would have experienced parallel traffic trends absent the tax and that the comparison group is unaffected. For the main synthetic comparison, I fit a separate donor vector for each of the 469 advertising-exposed nonliable families, drawing from the unexposed group. I apply the French-only fitting criterion described below, with French-only donor eligibility and no French-audience exposure balancing term. Figure 5 and Table 8 report the resulting paths and period estimates.

I also estimate a dynamic version of Equation 8:

$$y_{it} = \alpha_i + \lambda_t + \sum_{m \in \mathcal{T} \setminus \{\text{Nov. 2018}\}} \beta_m (L_i \times \mathbf{1}\{t = m\}) + \varepsilon_{it}, \quad (10)$$

where $\mathcal{T}$ contains the estimation months and November 2018 is the omitted reference month. Each coefficient β_m measures the liable–nonliable difference in log traffic in month m relative to that difference in November 2018. For presentation, I recenter the coefficients on their September–November 2018 average:

$$\tilde{\beta}_m = \beta_m - \frac{1}{3} \sum_{s \in \mathcal{P}} \beta_s, \quad \mathcal{P} = \{\text{Sep., Oct., Nov. 2018}\}, \quad (11)$$

with $\beta_{\text{Nov. 2018}} = 0$. The dynamic estimates show how the relative traffic response evolves during 2019 and whether differential changes are visible within the baseline period. Standard errors and pointwise percentile intervals use the full-history bootstrap described below.

A causal interpretation requires that, absent the tax announcement, liable and nonliable websites would have experienced parallel changes in log traffic. It also requires that the tax does not affect the comparison websites. This second condition deserves particular attention because nonliable websites can interact with liable platforms through advertising and other services. If these relationships transmit the tax’s effects, the estimate captures the response of liable websites relative to an affected comparison group.

I estimate the corresponding dynamic specification among nonliable website families by replacing (L_i) with the advertising-exposure indicator (E_i) . Its monthly coefficients measure exposed–unexposed differences relative to November 2018. For presentation, I recenter these coefficients on their September–November 2018 average, using the same normalization as above.

The conceptual framework in Section 3 identifies competing effects of the tax on advertising-supported platforms. A reduction in advertising intensity can increase traffic, while a reduction in user-facing expenditure can decrease it. I examine whether traffic responses vary with reliance on advertising using advertising’s share of in-scope revenue.

The specification measures advertising’s share of in-scope consolidated revenue, defined as advertising revenue divided by the sum of advertising and intermediation revenue:

$$y_{it} = \alpha_i + \lambda_t + \beta (L_i \times \text{Post}_t) + \theta_C \left(L_i \times \text{Post}_t \times \tilde{C}_{j(i)} \right) + \varepsilon_{it}, \quad (12)$$

where $j(i)$ denotes the parent firm of website family i and $C_j = R_j^A / (R_j^A + R_j^I)$. The variables R_j^A and R_j^I denote consolidated advertising and intermediation revenue as reported in Table 2. This ratio describes the composition of revenue from the two in-scope activities, rather than their importance in the firm’s total revenue.

I construct C_j from fiscal-year 2018 accounts and hold it fixed throughout the analysis. The standardized measure $\tilde{C}_j$ subtracts its mean and divides by its standard deviation among the 55 included liable website

families. The coefficient β compares liable and nonliable traffic changes at mean advertising composition; θ_C is the gradient per one-standard-deviation increase in composition. The estimation period, equal family-month weights, fixed effects and bootstrap resampling blocks follow the baseline specification. For synthetic comparisons, I use the same revenue interaction on the liable member and retain member and shared month fixed effects.

A causal interpretation of this gradient requires that, absent the tax announcement, traffic changes among liable websites would not have varied systematically with advertising composition.

I report both the coefficient estimates and the implied traffic responses for hypothetical pure advertising and pure intermediation firms. The coefficients describe how the response varies with revenue composition, while the hypothetical profiles express these relationships as predicted liable–nonliable differences in traffic changes.

For Equation 12, a pure advertising firm has $C_j = 1$ and a pure intermediation firm has $C_j = 0$. Their implied responses are

$$\Delta_A^C = \beta + \theta_C \frac{1 - \bar{C}}{s_C}, \quad (13)$$

$$\Delta_I^C = \beta + \theta_C \frac{-\bar{C}}{s_C}, \quad (14)$$

where $\bar{C}$ and s_C are the mean and standard deviation used to standardize C_j . Here, “pure” refers to the composition of in-scope revenue; either firm may also earn revenue from other activities.

For each hypothetical profile, I evaluate the full expression in every bootstrap draw. Standard errors are the standard deviation of the resulting estimates, retaining dependence between the intercept and composition slope; confidence intervals are their 2.5th and 97.5th percentiles.

4.1 Empirical Bayes estimates of average website effects.

The hypothetical profiles above describe predictions from the composition model. To estimate average effects among observed pure advertising and pure intermediation websites while allowing individual websites to depart from that relationship, I employ an Empirical Bayes shrinkage strategy. The calculation uses all 55 liable families and their unexposed synthetic comparisons. For each family, I first calculate the raw annual contrast

$$\hat{\delta}_i = \frac{1}{12} \sum_{t \in \mathcal{Q}} g_{it} - \frac{1}{3} \sum_{t \in \mathcal{P}} g_{it}, \quad \mathcal{Q} = \{\text{Jan.} - \text{Dec. 2019}\}, \quad (15)$$

where g_{it} is the observed-minus-synthetic log traffic gap and $\mathcal{P}$ contains September–November 2018. December 2018 is excluded. The composition model supplies the center

$$m_i = \hat{\beta} + \hat{\theta}_C \frac{C_{j(i)} - \bar{C}}{s_C}, \quad (16)$$

using the original stacked synthetic-control specification in Equation (12). This center is not separately refitted by regressing website gaps on composition.

The working empirical Bayes model is

$$\widehat{\delta}_i \mid \delta_i \sim N(\delta_i, s_i^2), \quad \delta_i \mid C_{j(i)} \sim N(m_i, \tau^2), \quad (17)$$

where s_i^2 is the estimated sampling variance of the raw contrast and τ^2 measures dispersion of website effects around the composition prediction. To calibrate s_i^2 , I fit each website’s monthly gap over the three baseline months and twelve months of 2019 with a website-specific baseline and a constant website-specific post-announcement shift. Residuals follow a common AR(1) correlation structure with website-specific variance scales. Contrast variances account for baseline covariance and actual calendar distances, including the gap between November and January. December is excluded from this calibration. This working noise model treats fluctuations around a website’s constant 2019 shift as noise; genuine dynamic responses or synthetic mismatch can therefore contribute to the estimated sampling variance.

Holding the fitted centers fixed during the variance optimization, I estimate the nonnegative residual variance τ^2 by maximizing the plug-in Gaussian marginal likelihood of $\widehat{\delta}_i$ with mean m_i and variance $\tau^2 + s_i^2$. The resulting website estimate is

$$\widetilde{\delta}_i = r_i \widehat{\delta}_i + (1 - r_i) m_i, \quad r_i = \frac{\widehat{\tau}^2}{\widehat{\tau}^2 + \widehat{s}_i^2}. \quad (18)$$

More precise raw contrasts receive greater weight; less precise contrasts shrink toward the composition model. I use r_i to distinguish this reliability weight from synthetic donor weights. I average $\widetilde{\delta}_i$ equally across the 13 observed pure advertising families ($C = 1$, five parents) and, separately, the five pure intermediation families ($C = 0$, five parents). These observed-group averages differ from the hypothetical model-only profiles Δ_A^C and Δ_F^C . All reported effects are multiplied by 100, without exponentiation.

Annual raw-group and EB standard errors and percentile confidence intervals use 1,999 full-pipeline bootstrap draws. Parent sampling is stratified by the three fixed revenue-composition groups: five advertising parents, ten mixed parents and five intermediation parents are sampled with replacement within their groups. Each selected parent’s entire set of website families and months is retained. The 169 residual nonliable family histories are sampled separately. Each draw reconstructs synthetic paths, reestimates the original stacked-model centers, AR(1) correlation, website noise scales, residual signal variance, reliability weights and group averages. Within a draw, the centers are treated as given only during residual-variance optimization; they are reestimated across draws. Inference therefore incorporates donor-sample and synthetic-weight variation as well as EB parameter estimation.

For the dynamic analysis, I apply the same construction separately to each month using the original monthly stacked-model predictions, month-specific raw gap contrasts, and sampling variances from the same AR(1) calibration. The residual signal variance is estimated separately for each month. I recenter fitted

website paths on their own September–November 2018 means and average equally within each reported group. December 2018 enters as a separate transition month, while remaining excluded from the annual contrast and noise calibration. Pointwise 95% percentile intervals use the same 1,999 revenue-stratified full-pipeline draws, repeating synthetic fitting, monthly signal estimation and normalization in every draw. These monthly fits do not impose a joint Bayesian dynamic model. The annual empirical Bayes estimate is constructed from annual contrasts and need not equal the average of the separately fitted monthly estimates.

Appendix F reports the specifications and results for the supplementary French-share analysis (Figure F.1 and Table F.1) and joint revenue-share analysis (Figure F.2 and Table F.2).

The following subsections describe sample construction and the characteristics of nonliable websites, including their connections to liable platforms.

4.2 Data and Sample Construction

With this estimation strategy in mind, I construct a monthly panel of website-family level data on French-origin web traffic as follows.

WhoTracksMe reports monthly website popularity relative to Google within each geographic market. The original measure therefore reflects a website’s traffic relative to Google’s traffic, rather than an absolute traffic count. Let p_{dFt} denote this Google-normalized popularity for domain d , among users in France, in month t .

I first aggregate domains into website families by summing their popularity:

$$P_{it}^F = \sum_{d \in \mathcal{D}_i} p_{dFt},$$

where $\mathcal{D}_i$ contains the domains belonging to family i . Because the member domains share the same Google denominator, their sum expresses family traffic relative to Google.

I then renormalize family popularity relative to Wikipedia by dividing by Wikipedia’s Google-normalized popularity in the same month and geographic market:

$$y_{it} = \log \left(\frac{P_{it}^F}{p_{\text{Wikipedia}, Ft}} \right).$$

The original Google denominator cancels in this ratio, leaving the logarithm of family traffic relative to Wikipedia.

The choice of reference website does not affect within-month treated–control log differences, because the common reference term also cancels:

$$y_{it} - y_{jt} = \log P_{it}^F - \log P_{jt}^F.$$

The same cancellation holds for a given synthetic comparison whose weights sum to one. In the difference-

in-differences regressions, month fixed effects absorb the common monthly reference term.

I restrict the panel to website families with sufficient observations before the December 2018 announcement to assess pre-treatment traffic patterns. For families owned by liable corporations, I generally require at least four months of observations between June and November 2018. I retain Futhead as an exception, with observations for September–November 2018. For nonliable families, I require observations in all six months. This stricter coverage requirement addresses the sharp expansion in the number of websites published by WTM during the pre-announcement period, which leaves many potential comparison websites with short traffic histories.

The resulting panel contains 55 liable website families belonging to 20 parent firms (Table 3) and 638 nonliable website families. I retain available observations during January–December 2019 without requiring complete post-announcement coverage. The estimation baseline is September–November 2018; December 2018 is excluded from the annual comparisons because the announcement occurs during that month.

The heterogeneity specifications attach parent-level revenue measures to each liable website family. The French audience concentration measure is instead constructed at the family level using pre-announcement French and global popularity. These characteristics are held fixed over time. The following subsections describe the advertising connections used to define comparison groups and the synthetic-control construction.

4.3 Advertising Exposure among Nonliable Websites

The advertising connections described in Subsection 2.4 can transmit the DST to nonliable websites. In this section I describe how I leverage historical data on the presence of advertising-related tracking technologies to connect third party sites to advertising activity related to the liable platforms in this study.

I start with the WhoTracksMe dataset, which records detections of web technologies on individual website domains in its country-specific monthly panels. These technologies include advertising services, analytics tools, and other website infrastructure. For each detected technology, WTM provides an identity and category classification. I use detections in its `advertising` category in the French panel to measure observed advertising connections among nonliable websites.

Table 6 reports observed advertising connections among the 638 nonliable website families with balanced pre-treatment traffic observations. Exposure requires a detection on at least one member domain during September–November 2018. The table distinguishes connections to any advertising provider from connections to providers owned by liable corporations.

These measures identify technical connections rather than advertising expenditures, receipts, or contractual relationships. An observed connection does not establish whether a website buys or sells advertising, how much revenue is involved, or whether the tax is passed through to it. Conversely, the absence of a qualifying observation does not establish the absence of advertising relationships. I therefore label the residual category “no observed exposure,” including families whose exposure is unknown.

This exercise identifies 169 website families without an observed advertising connection to liable plat-

forms. These families form the starting pool from which eligible synthetic-control donors are selected. The remaining 469 families have observed connections to liable advertising providers; I analyze their traffic separately to estimate spillover effects on nonliable websites.

Table 6: Observed advertising exposure among nonliable website families
September–November 2018

Exposure measure	Observed	Not observed
Any-provider advertising	531	107
Liable advertising	469	169

Notes: The table classifies 638 nonliable website families using positive published detections in the French WTM panel and identities in its `advertising` category. A detection on any member domain in any month of September–November 2018 identifies exposure for the family. Any-provider advertising includes providers regardless of liability; liable advertising includes only providers owned by firms classified as liable in this study. The main measure is the liable-provider row: 469 observed and 169 not observed. Rows overlap and should not be added. Categories use the saved April 2026 metadata and liability uses the study crosswalk.

4.4 Constructing a Panel of Synthetic Controls

My preferred estimation strategy uses stacked synthetic controls. For each of the 55 liable website families and 469 advertising-exposed nonliable families, I construct a synthetic comparison from nonliable families without observed pre-treatment advertising connections to liable platforms.

From the 169 eligible sites for whom I do not observe liable-tech exposure, I further impose full balance for the entire analysis period. This results in 159 donors.

For each target family i , the French synthetic path is

$$\hat{y}_{it}^0 = a_i + \sum_{j \in \mathcal{D}_i} w_{ij} y_{jt}, \quad w_{ij} \geq 0, \quad \sum_{j \in \mathcal{D}_i} w_{ij} = 1. \quad (19)$$

I demean each French target and donor history over the target’s observed June–November 2018 training months. The mean squared trajectory loss is divided by the variance of donor traffic changes over those months, and a ridge penalty with coefficient $\sqrt{12}$ limits weight concentration. Weights are nonnegative and sum to one. I choose a_i so that the synthetic path matches the target’s average French traffic over September–November 2018.

Let $\mathcal{T}_i$ denote the target’s observed June–November 2018 months, and let $\tilde{y}_{kt}^{(i)}$ demean each history over those same months. Then:

$$\hat{\mathbf{w}}_i = \arg \min_{\mathbf{w}_i \geq 0, \mathbf{1}'\mathbf{w}_i=1} \left\{ \frac{1}{|\mathcal{T}_i| \hat{\sigma}_i^2} \sum_{t \in \mathcal{T}_i} \left(\tilde{y}_{it}^{(i)} - \sum_{j \in \mathcal{D}_i} w_{ij} \tilde{y}_{jt}^{(i)} \right)^2 + \sqrt{12} \sum_{j \in \mathcal{D}_i} w_{ij}^2 \right\}.$$

Here, $\hat{\sigma}_i^2$ is the sample variance of donor month-to-month traffic changes, pooled across donors and consecutive months observed for the target. The first term rewards matching the pre-announcement trajectory; the

second discourages concentrating weight on a few donors.

To estimate the weights, I collect pre-announcement traffic histories into a feature vector $\mathbf{x}_i(\mathcal{F})$, where $\mathcal{F}$ denotes the set of countries whose traffic histories enter the fitting criterion. I consider French traffic alone and French traffic combined with German and US traffic. For each included country, I use the target's observed training months and demean each target and donor history over those same months. I divide each country's demeaned histories by $\sqrt{n_{ic}\hat{\sigma}_{ic}}$, where n_{ic} is the number of observed training months for target i in country c , and $\hat{\sigma}_{ic}$ is the standard deviation of donor traffic changes, pooled across eligible donors and consecutive observed training months. This scaling expresses discrepancies relative to donor traffic fluctuations and prevents countries with more observed months from mechanically receiving greater weight.

Let $\mathbf{x}_j^{(i)}(\mathcal{F})$ denote the corresponding feature vector for donor j , constructed using target i 's training months and scaling factors. I estimate weights by solving

$$\hat{\mathbf{w}}_i(\mathcal{F}) = \arg \min_{\substack{w_{ij} \geq 0 \\ \sum_{j \in \mathcal{D}_i(\mathcal{F})} w_{ij} = 1}} \left\{ \left\| \mathbf{x}_i(\mathcal{F}) - \sum_{j \in \mathcal{D}_i(\mathcal{F})} w_{ij} \mathbf{x}_j^{(i)}(\mathcal{F}) \right\|_2^2 + \lambda \sum_{j \in \mathcal{D}_i(\mathcal{F})} w_{ij}^2 \right\}. \quad (20)$$

The first term rewards matching the target's pre-announcement traffic trajectory. The second discourages concentrated weights. I hold the regularization coefficient fixed at $\lambda = \sqrt{12}$ across specifications; the validation exercise below selects between fitting procedures rather than tuning this coefficient. Once the weights are estimated using all available June–November 2018 training observations, I choose the intercept to align the synthetic path with the target's September–November 2018 mean:

$$\hat{a}_i = \frac{1}{3} \sum_{t \in \{\text{Sep., Oct., Nov. 2018}\}} \left(y_{it} - \sum_{j \in \mathcal{D}_i} \hat{w}_{ij} y_{jt} \right). \quad (21)$$

I assess the value of adding foreign traffic histories using three forward validation splits within the pre-announcement period. The first fits June–August 2018 and predicts September–November; the second fits June–September and predicts October–November; the third fits June–October and predicts November. The evaluation samples contain 49, 54 and 55 liable website families, respectively. Within each split, I estimate all demeaning transformations, donor-change scales and weights using training observations only. For prediction, I estimate the French intercept from the mean observed-minus-weighted-donor gap in the training months, without recentering on held-out observations.

For each target and split, I calculate the root mean squared prediction error over its observed held-out French months. I average these errors equally across target families within each split, then give equal weight to the three split averages. French-only fitting uses the 159 donors with complete French coverage. The multinational procedure uses smaller, target-specific donor sets satisfying additional German, US and global coverage requirements, although global traffic does not enter its fitting criterion. The exercise therefore

compares the predictive performance of these two procedures, including their donor eligibility rules.⁵

I retain French-only fitting unless the multinational procedure reduces the average held-out French prediction error by at least two bootstrap standard errors. I estimate the standard error of this difference using 9,999 paired whole-parent bootstrap draws, resampling the 20 liable parent firms and retaining all website families belonging to each sampled parent. Each draw uses the same parent multiplicities across procedures and validation splits and recomputes the average prediction-error difference from the saved held-out errors. This bootstrap conditions on the fitted predictions; it does not refit the synthetic weights. For the main donor group, the multinational procedure improves prediction error by 0.076 log percentage points, with a bootstrap standard error of 0.482. This improvement falls below the two-standard-error threshold, so I use French-only fitting.

4.5 Inference for the full estimation pipeline

For the main regression tables and monthly paths, I use 1,999 pairs-bootstrap draws. Each draw samples 20 liable parents with replacement, retaining every family and month under a selected parent. I separately sample 469 exposed nonliable families and 169 residual nonliable families with replacement, retaining their full histories and observed missingness. The residual, exposed and pooled donor sets are selected from these sampled family copies using the original French completeness rule.

In each draw, I recompute donor-change scales and demeaned pre-announcement histories, fit the non-negative unit-sum synthetic weights with $\lambda = \sqrt{12}$, realign the September–November baseline, and reestimate the annual and monthly regressions. Duplicate donor copies are separate coordinates in the simplex and each receives its own ridge penalty, as in the expanded resampled dataset. Available observations and original equal family-month weights are retained. Composition is standardized using the original 55-family mean and SD throughout, preserving the coefficient’s unit. Actual-control regressions use the same sampled histories. Standard errors are the sample standard deviation across draws; plotted 95% intervals use the 2.5th and 97.5th percentiles and are pointwise rather than simultaneous.

Model-only profiles use these ordinary parent draws. The observed pure-group raw and EB means instead use the revenue-stratified parent draws described above, repeating the same donor resampling and synthetic fitting. December 2018 is excluded from annual contrasts and primary monthly regressions. In Figures 4–5, it appears only as a separately estimated transition point without an interval. The monthly EB calculation retains its original separate December transition in nuisance fitting, while Figure 6 omits it; the annual noise calibration excludes December.

Appendix H reports an alternative approach that holds fitted synthetic controls fixed when calculating standard errors and confidence intervals.

⁵Donor membership, coverage requirements and advertising-exposure classifications are held fixed across validation splits, rather than reconstructed using only information available at each training date. No held-out traffic values enter the fitting criterion.

4.6 Results

Traffic responses among liable sites Figure 4 presents monthly French traffic paths and liable-minus-comparison contrasts for the 55 liable website families in my sample. The left column uses individual synthetic controls fitted to prepolicy French traffic only; the right column uses equal-weighted actual controls. Both columns use the comparison group without documented advertising exposure. The upper panels display the liable and comparison paths, and the lower panels display their differences. Table 7 reports the corresponding period DiD coefficients and bootstrap standard errors for actual and synthetic comparisons, using the unexposed control group. Appendix G, Table G.1, reports the corresponding comparisons using advertising-exposed and pooled controls. The table also reports sample counts, fixed effects and regression fit statistics.

The stacked synthetic-control estimate is the preferred specification, using French-only prepolicy fitting and the residual advertising-exposure group. It indicates a traffic decline of 2.81 log percentage points (s.e. 7.78) among liable platforms relative to their synthetic comparisons. Scaling by the 3-percentage-point statutory DST gives a response of -0.94 log percentage points of traffic per percentage-point increase in the statutory rate.

Table 7: Difference-in-differences estimates of French traffic responses

Model Estimator Dependent variable	No documented advertising exposure	
	(1) DiD y_{it}	(2) Stacked SC y_{ist}
Liable $\times$ Post (β)	+5.74 (7.56)	−2.81 (7.78)
Liable families	55	55
Liable parent firms	20	20
Raw comparison families	169	–
Synthetic comparison units	–	55
Eligible donor pool	–	159
Eligible donors per liable family	–	159
Regression units	224	110
Observations	3,311	1,650
Liable / nonliable resampling blocks	20 / 169	20 / 169
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
R^2	0.953	0.994
Within R^2	0.00073	0.00101
RMSE (log units)	0.369	0.177

Notes: January–December 2019 versus September–November 2018; December 2018 excluded. Coefficients are 100β , with standard errors from 1,999 full-history bootstrap draws in parentheses. Column (1) estimates Equation (8) on observed families; column (2) stacks each liable family with its synthetic member, retaining member and shared month fixed effects. Each draw samples 20 whole liable parents and, separately, 169 residual nonliable families with replacement. All family histories and missingness remain together. Eligible donor copies are selected using French coverage; donor-change scales, synthetic weights and baseline offsets are reestimated before refitting the regression.

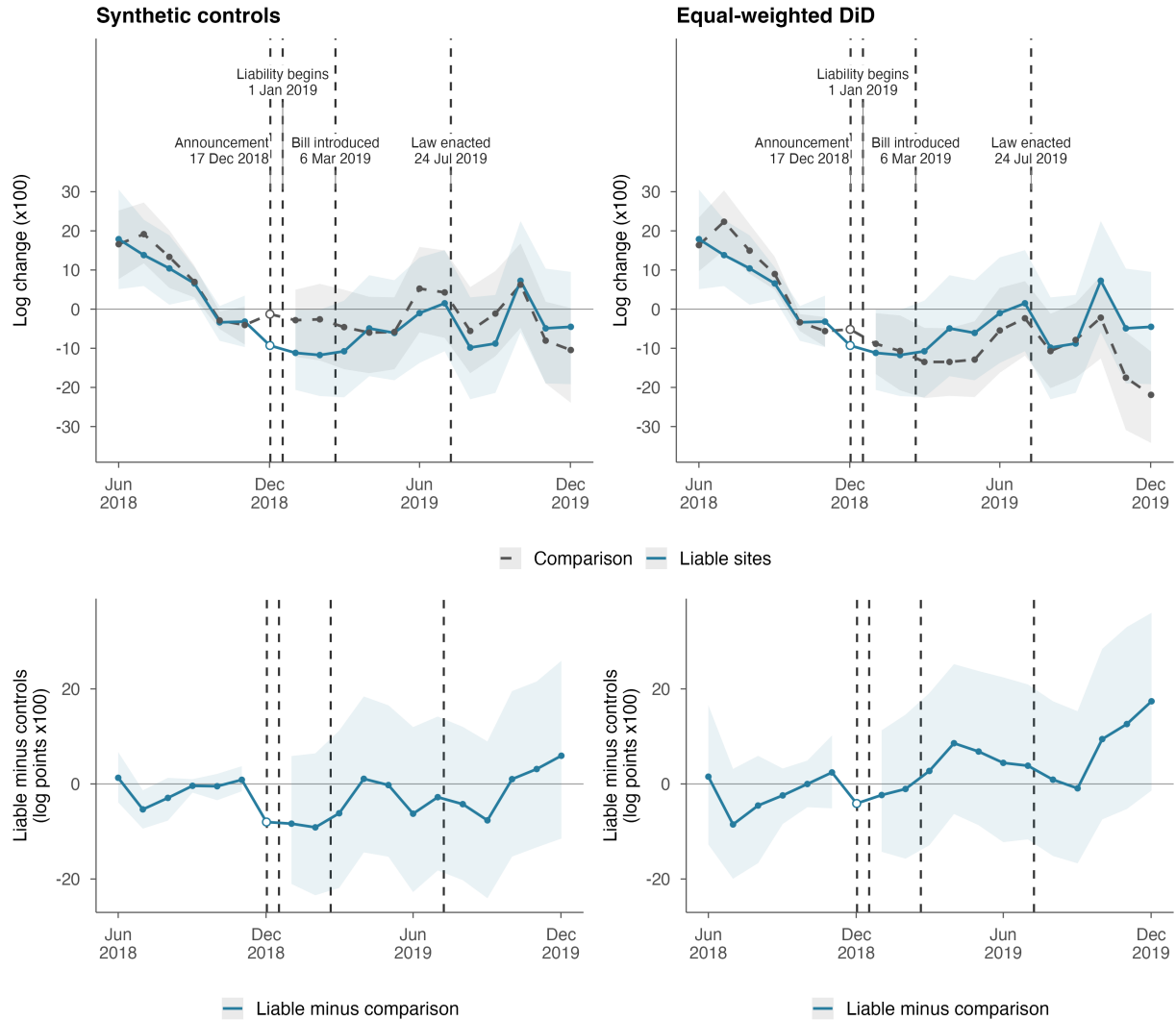


Figure 4: French traffic responses of liable platforms

Notes: Left: synthetic controls; right: equal-weighted actual-control DiD. Both use 55 liable families under 20 parents and the unexposed nonliable group (169 families; 159 eligible synthetic donors). Upper panels show liable and comparison paths; lower panels show their differences. Series are centered on September–November 2018 and displayed as 100 times log changes or differences. Shading gives pointwise 95% percentile intervals from 1,999 draws of whole liable parents and residual nonliable family histories. Synthetic scaling, weights, baseline offsets and regressions are reestimated in every draw. December 2018 is display-only and has no interval.

Traffic responses of advertising-exposed nonliable sites. Figure 5 compares 469 advertising-exposed nonliable families with unexposed families, using synthetic controls drawn from the unexposed donor pool. Its left column uses individual synthetic controls and its right column uses equal-weighted actual unexposed controls. The upper row displays the exposed and comparison paths; the lower row displays the corresponding exposed-minus-comparison differences. Table 8 reports the exposed-by-post coefficient and bootstrap standard error for equal-weighted actual-control DiD and stacked synthetic DiD, together with sample counts, fixed effects and fit statistics.

The stacked synthetic comparison estimates a traffic decline of 8.24 log percentage points (s.e. 4.96) among advertising-exposed nonliable families. This pattern is consistent with effects transmitted through advertising relationships. Remember, however, that in practice this may include not only reductions in advertising on the liable platforms but also the effect of diminished returns on ads displayed on behalf of liable corporations. The exposure measure can identify websites that buy advertising, display advertising or use associated tracking and measurement technologies; it does not separate the roles or establish which channel generates the traffic decline.

Table 8: Difference-in-differences estimates for advertising-exposed nonliable sites

Model	(1)	(2)
Estimator	Conventional DiD	Stacked synthetic DiD
Dependent variable	y_{it}	y_{ist}
Exposed $\times$ Post (δ)	-1.75 (5.30)	-8.24 (4.96)
Exposed nonliable families	469	469
Raw unexposed families	169	—
Synthetic comparison units	—	469
Starting eligible donor pool	—	159
Eligible donors per exposed family	—	159
Regression units	638	938
Observations	9,506	14,040
Exposed / residual resampling blocks	469 / 169	469 / 169
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
R^2	0.895	0.934
Within R^2	0.00004	0.00212
RMSE (log units)	0.477	0.358

Notes: January–December 2019 versus September–November 2018; December 2018 excluded. Coefficients are 100δ , with bootstrap standard errors in parentheses. Column (1) estimates Equation (9) on actual nonliable families; column (2) stacks each exposed family with its own synthetic comparison. In each of 1,999 draws, 469 exposed and 169 residual nonliable families are sampled separately with replacement, keeping every selected family’s full history and missingness.

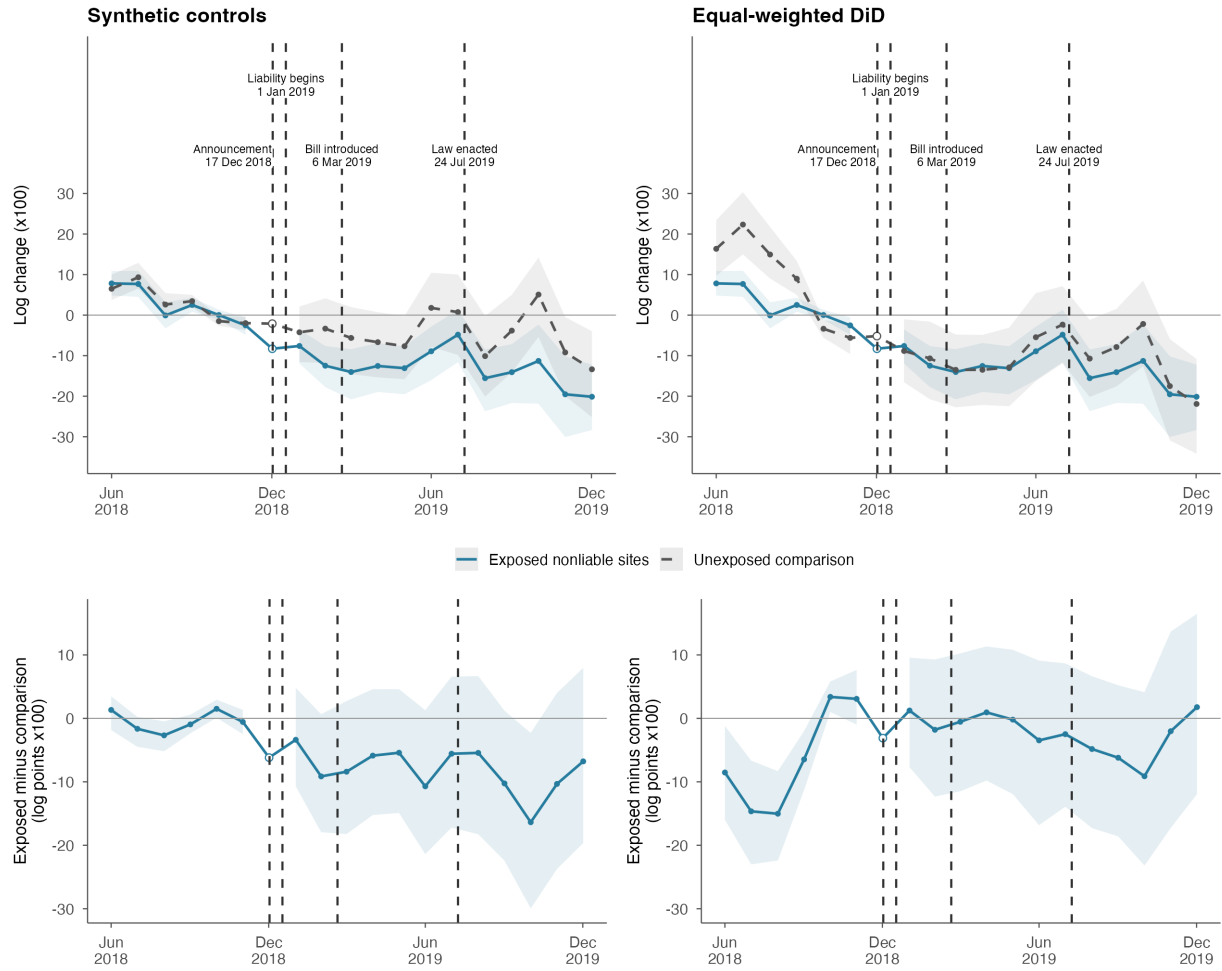


Figure 5: French traffic responses of advertising-exposed nonliable sites

Notes: Synthetic controls (left) and equal-weighted actual controls (right) compare 469 exposed nonliable families with 169 residual families; the synthetic pool contains 159 French-complete donors. Upper panels show normalized traffic paths; lower panels show exposed-minus-comparison contrasts. All series are centered on September–November 2018 and measured as 100 times log changes or differences. Shading gives pointwise 95% percentile intervals from 1,999 draws of full family histories within exposure groups. Every draw repeats synthetic scaling, weight estimation, baseline alignment and the regressions, sharing donor multiplicities across target comparisons.

4.7 Traffic responses by business model

Advertising composition of in-scope revenue. Table 9 reports β and θ_C from Equation (12), using actual-control and stacked synthetic DiD for the unexposed control group. Appendix Table G.2 reports the same composition specification using advertising-exposed and pooled controls. The table includes bootstrap standard errors, fit statistics, sample counts and fixed-effect indicators. The period regressions compare September–November 2018 with January–December 2019, excluding December 2018. Coefficients are displayed as 100 times log units; the composition gradient is per original liable-sample standard deviation.

Across both DiD and stacked synthetic specifications, a one-standard-deviation increase in advertising’s share of in-scope revenue is associated with a 12.38 log percentage point increase in traffic (s.e. 4.92).

To understand this estimate in the context of specific business models, Table 10 reports the corresponding annual pre/post effects from the composition model, raw pure-group averages and empirical Bayes estimates. Appendix E reports the annual raw synthetic-control contrast, composition-model prediction, empirical Bayes estimate, and weight on the raw contrast for each of the 55 liable website families. The hypothetical advertising endpoint is +16.48 (s.e. 9.11), and the intermediation endpoint is −17.50 (s.e. 10.09). Among observed pure-type families, the raw advertising and intermediation averages are +22.25 (s.e. 7.97) and +5.02 (s.e. 6.59). The corresponding EB averages are +16.90 (s.e. 7.81) and −11.13 (s.e. 8.24).

Figure 6, shows the dynamic counterparts of these estimates. They compare monthly composition-model predictions with raw and empirical Bayes averages among observed pure-type website families, as in websites whose parents report only advertising or only intermediation in-scope revenue.

Taken together, these results show that the modest and imprecise aggregate estimates presented in Table 7 conceal substantial heterogeneity by parent business model. Moreover, the estimated effects comport with the predictions of the models presented in Section 3. I discuss these results more closely in the following section.

5 Discussion

The empirical results are consistent with the key idea from Section 3, which is that although France’s DST framework places advertising and intermediation under the same digital service umbrella, those two services have fundamentally different relationships with user participation that are important to take into account when assessing the impact of DSTs on users. In practice, the empirical results suggest declines in traffic for pure intermediation sites and increases for pure advertising sites. Moreover, had I not taken this source of heterogeneity into account, I would have only obtained an incomplete understanding of the complex effects of web traffic on user behavior.

Seen through the lens of the models, I can map these traffic responses to unobservable platform decisions.

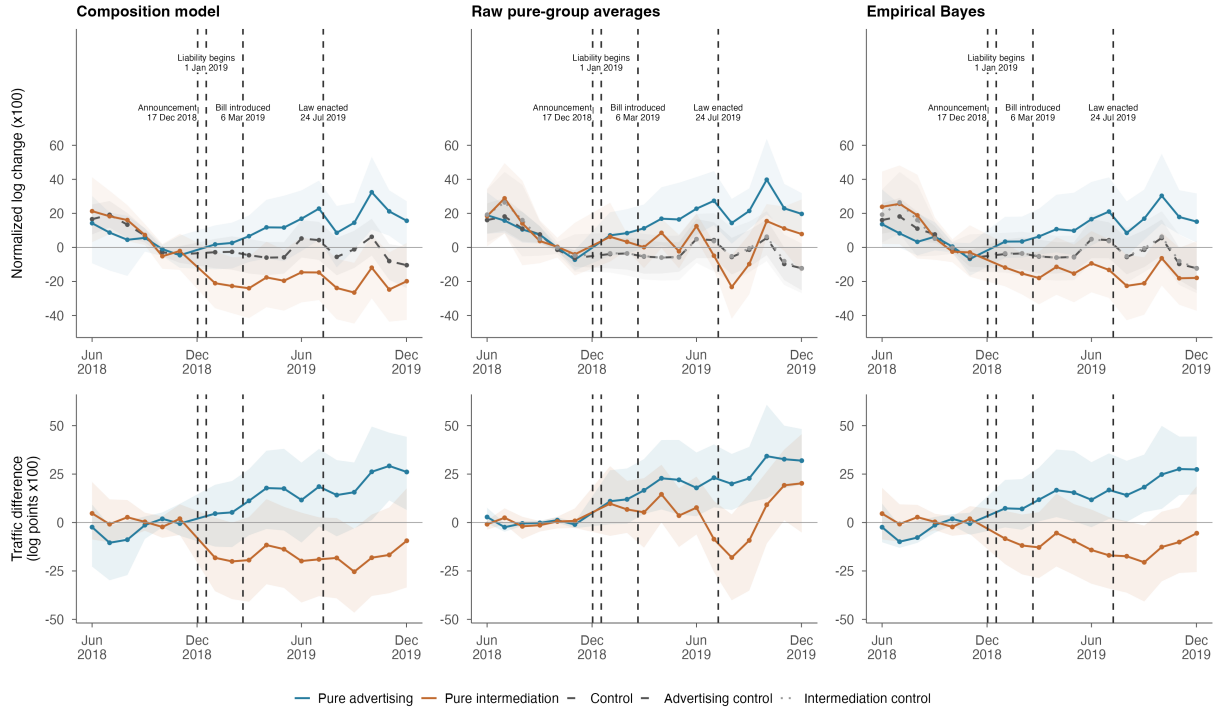


Figure 6: Advertising composition and traffic

Notes: Left: hypothetical pure advertising ($C = 1$, blue) and intermediation ($C = 0$, orange) profiles from the original stacked composition model (55 families, 20 parents). Middle: raw means among 13 observed pure advertising and 5 pure intermediation families. Right: EB means for the same groups, centered on the original model. Upper panels show normalized traffic paths and synthetic comparisons; lower panels show gaps. Paths are centered on September–November 2018 in 100 times log units; December is omitted. Shading gives pointwise 95% percentile intervals from 1,999 draws, with synthetic scaling, weights, baseline offsets and model coefficients refitted. The model uses ordinary whole-parent draws; raw and EB columns use revenue-stratified parent draws. Each draw samples residual nonliable donor families with shared multiplicities across targets. EB additionally refits noise correlation, website scales, monthly signal variances and reliability weights before recentering.

Table 9: Traffic responses by advertising composition of in-scope revenue

Estimator Model	No documented advertising exposure	
	Actual DiD (1)	Stacked SC (2)
Liabile $\times$ Post (β)	+5.74 (6.72)	−2.81 (6.99)
Liabile $\times$ Post $\times$ Composition (θ_C)	+12.38 (4.92)	+12.38 (4.92)
R^2	0.953	0.994
Within R^2	0.00513	0.03948
RMSE (log units)	0.368	0.173
Liabile families / parents	55 / 20	55 / 20
Raw comparison families	169	—
Synthetic comparison units	—	55
Observations	3,311	1,650
Regression units	224	110
Liabile / nonliable resampling blocks	20 / 169	20 / 169
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes

Notes: Equation (12); September–November 2018 versus January–December 2019, excluding December 2018. Coefficients and standard errors are multiplied by 100. Composition retains its original 55-family mean and SD, so β is the change at original mean composition and θ_C is the gradient per original-sample SD. Standard errors use 1,999 draws of 20 whole liable parents and 169 residual nonliable families. Every draw reestimates synthetic scaling, weights, baseline offsets and both regression coefficients; actual-control columns use the same resampled histories. The synthetic comparison changes the common intercept, while the gradient is determined by liable outcomes and is unchanged by refitting weights within a draw. Counts and fit statistics describe the original sample.

Table 10: Pre/post traffic effects by business model and estimation method

	Composition model	Raw group average	Empirical Bayes
Pure advertising	+16.48 (9.11)	+22.25 (7.97)	+16.90 (7.81)
Pure intermediation	−17.50 (10.09)	+5.02 (6.59)	−11.13 (8.24)

Notes: Annual effects in log percentage points, with bootstrap standard errors in parentheses. September–November 2018 versus January–December 2019; December 2018 excluded. The composition model evaluates the original stacked specification at $C = 1$ and $C = 0$, using 55 families under 20 parents. Raw and EB columns average equally over 13 observed pure advertising families or 5 pure intermediation families; EB centers use the original composition model. Each column uses 1,999 draws with synthetic donor weights refitted. The model column uses ordinary whole-parent draws; raw and EB columns use 5 advertising, 10 mixed and 5 intermediation parent draws within fixed revenue groups. All draws independently sample the 169 residual nonliable families, retaining complete histories. Scaling, weights, baseline offsets and model coefficients are reestimated; EB additionally refits AR(1) correlation, website noise scales, residual signal variance and reliability weights. Annual EB is calculated directly from annual contrasts.

I begin with advertising-funded platforms. The empirical Bayes average among observed pure advertising websites is an increase of 16.9 log percentage points (s.e. 7.8), while Figure 6 displays the composition model’s hypothetical pure advertising path. Under the maintained model assumptions, a positive traffic effect requires the benefit of reduced advertising intensity to dominate the loss from lower content and service provision in Equation (6).

I can make further progress on the advertising and content decomposition by using the estimated 8.24 log percentage point traffic decline at advertising-exposed nonliable websites. The calibration treats this decline as a proportional measure of the change in advertising impressions, holding click-through rates and the conversion of clicks into measured visits fixed and abstracting from other sources of traffic change. Translating a change in cost per click into the model’s price per impression also requires a stable click-through rate. These are additional measurement assumptions; Appendix C derives demand for impressions but does not establish these empirical mappings. Combining this quantity proxy with Acun et al. (2026)’s estimated 3.4 percent increase in effective costs per click, approximated as 0.034 log points, gives an illustrative advertising-demand elasticity:

$$\varepsilon_{A,p} \simeq \frac{\Delta \log A}{\Delta \log p} = \frac{-0.0824}{0.034} \simeq -2.42. \quad (22)$$

For a website-specific illustration, I use Google’s empirical Bayes traffic effect of +3.91 log percentage points. Applying the same advertising quantity response to Google additionally assumes that the pooled advertiser-site estimate and the external price evidence describe the relevant change in Google’s advertising market. Since $a = A/L$:

$$\Delta \log a = \Delta \log A - \Delta \log L = -0.0824 - 0.0391 = -0.1215. \quad (23)$$

Log linearizing Equation (6), and using the unit elasticity of traffic with respect to advertising intensity implied by the expenditure-minimization condition in Appendix C.2, yields an expression that allows me to estimate the content effect:

$$\Delta \log L \approx -\Delta \log a + \underbrace{\varepsilon_{L,g_U} \Delta \log g_U}_{\text{content/services effect}}, \quad (24)$$

Lower advertising intensity therefore contributes approximately +12.15 log percentage points to Google’s traffic, while its net traffic effect is +3.91. The implied residual contribution from content and services is −8.24 log percentage points.

The intermediation point estimates are consistent with the model’s negative traffic prediction. Langenmayr and Muddasani (2025) report a French consumer log-price effect of 0.022, or 2.2 log percentage points, in Table 5; their French seller-count estimate in Table A.4 is imprecise. Combining that price response with Amazon’s empirical Bayes traffic decline of 4.50 log percentage points gives an illustrative upper bound on

the magnitude of the traffic-price elasticity when both higher prices and reduced seller availability weakly reduce traffic:

$$0 \leq |\varepsilon_{L,p_B}| \leq \frac{-\Delta \log L}{\Delta \log p_B} \simeq \frac{0.0450}{0.022} = 2.05. \quad (25)$$

This estimate is a touch larger than prominent e-commerce price estimates of demand in the Economics and Finance literature, where [Houde et al. \(2023\)](#) estimates e-commerce price elasticities of 1.52, [Einav et al. \(2014\)](#) estimates -1.67, and [Baugh et al. \(2018\)](#) estimates -1.2 to -1.4. Assuming that demand elasticities appear proportionately in traffic, this implies that one quarter to one third of the estimated traffic response reflects declines in product availability or other declines in platform quality. The exercise illustrates how traffic can complement price evidence while making clear the additional assumptions needed to distinguish the channels.

6 Conclusion

Digital services taxes change the return to attracting and serving users, with consequences that depend on how platforms earn revenue. This paper studies those consequences around the December 2018 announcement of France’s DST, using web traffic to compare user participation across platforms with different business models. Relative to synthetic controls without documented advertising exposure, I estimate an average traffic decline of 2.8 log percentage points (s.e. 7.8) across the 55 liable website families. This average is imprecisely estimated and masks differences across business models. Empirical Bayes estimates indicate a decline of 11.1 log percentage points (s.e. 8.2) for pure intermediation websites and an increase of 16.9 (s.e. 7.8) for pure advertising websites.

The theoretical framework explains how the same turnover tax can generate these contrasting responses. For transaction intermediaries, higher buyer prices and reduced seller supply can both discourage visits. For advertising-supported platforms, lower expenditure on user services discourages participation, while reduced advertising intensity can make the platform more attractive. Within the model, a positive traffic response implies that the benefit of fewer advertisements outweighs the loss from reduced services. The estimates are consistent with these mechanisms, although traffic alone does not separately identify the underlying platform choices.

The observed changes also extend beyond websites owned by liable firms. Relative to synthetic controls without documented advertising exposure, advertising-exposed nonliable websites show an estimated traffic decline of 8.2 log percentage points (s.e. 5.0). This point estimate is consistent with the transmission of tax effects through advertising relationships and highlights the distinction between statutory liability and economic exposure. It also helps explain why estimated responses among liable platforms depend on the advertising connections of the comparison group.

The broader contribution is to make user participation part of the assessment of digital taxation, including on services that users access without paying a monetary price. Traffic captures responses to changes in the services, advertising, and transaction terms users encounter, but does not by itself measure welfare. In particular, an advertising tax can increase platform use even while reducing the activity it taxes, complicating proposals that also seek to discourage excessive use. Evaluating digital services taxes therefore requires attention to how platforms monetize participation, how their operations adjust, and how those adjustments reach users and businesses beyond the statutory tax base.

References

- Acemoglu, D. and Johnson, S. (2024). The urgent need to tax digital advertising. *Network Law Review*, Spring, 5.
- Acun, I., Kasinger, J., and Abou Nabout, N. (2026). Who bears the burden of digital services taxes? *Available at SSRN 6873483*.
- Adegbola, A. (2026). Meta is passing europe’s digital taxes directly to advertisers. Search Engine Land. Reports Meta’s advertiser notice: France location fee of 3 percent effective July 1, 2026.
- Adevinta (2023). Annual report 2022. <https://adevinta.com/wp-content/uploads/2023/04/Adevinta-AR2022.pdf>. Note 21.3.1, p. 147.
- Agrawal, D. R. and Fox, W. F. (2017). Taxes in an e-commerce generation. *International Tax and Public Finance*, 24(5):903–926.
- Agrawal, D. R. and Fox, W. F. (2021). Taxing goods and services in a digital era. *National tax journal*, 74(1):257–301.
- Amazon (2018). Amazon Publisher Services launches transparent ad marketplace in europe. <https://press.aboutamazon.com/uk/2018/1/amazon-publisher-services-launches-transparent-ad-marketplace-in-europe>. January 9.
- Amazon (2019). Testimony at the section 301 hearing on france’s digital services tax. Office of the United States Trade Representative, hearing transcript. Printed p. 45, lines 13–16; Amazon’s announced seller-fee increase effective October 1, 2019.
- ANGI Homeservices (2020). Form 10-k for the year ended december 31, 2019. <https://www.sec.gov/Archives/edgar/data/1705110/000170511020000003/angi-20193112x10k.htm>. Management discussion and analysis.
- Apple (2020). Upcoming tax and price changes for apps and in-app purchases. Apple Developer News.
- Assemblée nationale (2018). Compte rendu de la deuxième séance du lundi 17 décembre 2018.
- Assemblée Nationale (2023). Rendement de la taxe sur les services numériques. Journal Officiel. Written question no. 2974 (C. Sitzenstuhl); response from the Ministère de l’Économie published in the Journal Officiel, 14 February 2023, p. 1448. Firm counts sourced to the DGFIP.
- Baugh, B., Ben-David, I., and Park, H. (2018). Can taxes shape an industry? evidence from the implementation of the “Amazon tax”. *The Journal of Finance*, 73(4):1819–1855.

- Bibler, A. J., Teltser, K. F., and Tremblay, M. J. (2021). Inferring tax compliance from pass-through: Evidence from airbnb tax enforcement agreements. *Review of Economics and Statistics*, 103(4):636–651.
- Booking Holdings (2020). Form 10-k for the year ended december 31, 2019. <https://www.sec.gov/Archives/edgar/data/1075531/000107553120000011/bkng1231201910k.htm>. Management discussion and analysis.
- Bureau van Dijk, a Moody’s Analytics company (2025). Orbis. Global company database covering financials, ownership structures, and subsidiaries.
- Council of the European Union (2019). Outcome of the economic and financial affairs council meeting of 12 march 2019. Press release, Brussels.
- Criteo (2019). 2018 annual report. https://filecache.investorroom.com/mr5ir_criteo/669/download/CRTO_Annual%20Report%202018.pdf. FY2018 report, pp. 15–16.
- DGFIP (2021). Taxe sur les services numériques: Obligations déclaratives et comptables, régime général. <https://bofip.impots.gouv.fr/bofip/12004-PGP.html/identifiant=BOI-TCA-TSN-30-10-20210408>. BOI-TCA-TSN-30-10, paragraphs 190–200, version of April 8, 2021.
- Direction générale des Finances publiques (2019). Présentation de la taxe sur les services fournis par les grandes entreprises du secteur numérique.
- Einav, L., Knoepfle, D., Levin, J., and Sundaresan, N. (2014). Sales taxes and Internet commerce. *American Economic Review*, 104(1):1–26.
- European Commission (2018). Proposal for a council directive on the common system of a digital services tax on revenues resulting from the provision of certain digital services. Technical Report COM(2018) 148 final, European Commission, Brussels.
- Facebook (2018). Hard questions: What information do Facebook advertisers know about me? <https://about.fb.com/news/2018/04/data-and-advertising/>. April 23; by Rob Goldman, Vice President, Ads.
- Google (2021). Jurisdiction-specific surcharges. Google Ads Help. France regulatory operating cost effective May 1, 2021.
- Hines, J. R. (2023). Digital tax arithmetic. *National Tax Journal*, 76(1):119–143.
- Houde, J.-F., Newberry, P., and Seim, K. (2023). Nexus tax laws and economies of density in e-commerce: A study of Amazon’s fulfillment center network. *Econometrica*, 91(1):147–190.

- IAC/InterActiveCorp (2020). Form 10-k for the year ended december 31, 2019. <https://www.sec.gov/Archives/edgar/data/891103/000089110320000004/iac-20191231x10k.htm>. Management discussion and analysis.
- Karaj, A., Macbeth, S., Berson, R., and Pujol, J. M. (2018). Whotracks.me: Shedding light on the opaque world of online tracking. <https://arxiv.org/abs/1804.08959>.
- Kiin Labs (2026). How much do LinkedIn ads cost? 2026 benchmarks. <https://kiin.co/research/linkedin-ads-cost>. Accessed October 19, 2026; 2025–2026 benchmark period.
- Kind, H. J. and Schjelderup, G. (2025). Taxation and multi-sided platforms: a review. *International Tax and Public Finance*, 32(3):895–915.
- Langenmayr, D. and Muddasani, R. R. (2025). Navigating the amazon: The incidence of digital service taxes.
- Microsoft (2026). Universal event tracking: Microsoft Advertising API. <https://learn.microsoft.com/en-us/advertising/guides/universal-event-tracking?view=bingads-13>. Updated March 11, 2026; accessed October 19, 2026.
- République française (2019). Loi n° 2019-759 du 24 juillet 2019 portant création d’une taxe sur les services numériques. Journal officiel de la République française.
- Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029.
- Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: a progress report. *The RAND journal of economics*, 37(3):645–667.
- Romer, P. (2019). A tax that could fix big tech. *International New York Times*, pages NA–NA.
- Rysman, M. (2009). The economics of two-sided markets. *Journal of economic perspectives*, 23(3):125–143.
- Salesforce Marketing Cloud (2016a). Salesforce advertising index: Annual report 2015. <https://a.sfstatic.com/content/dam/www/ocms/assets/pdf/datasheets/q4-2015-advertising-index-salesforcemarketingcloud.pdf>. March; Q4 2015 price observations.
- Salesforce Marketing Cloud (2016b). Salesforce advertising index: Q2 2016 report. <https://brandcdn.exacttarget.com/sites/exacttarget/files/salesforce-advertising-index-q2-2016-advertising-studio-v8.pdf>. Q2 2016 price observations.
- Shapiro, C. and Varian, H. R. (1999). *Information rules: A strategic guide to the network economy*. Harvard Business Press.

- S&P Global Market Intelligence (2025). Compustat. Standardized financial statements for publicly traded firms; accessed via Wharton Research Data Services (WRDS).
- Statista (2026a). Facebook users in france and worldwide: Key market indicators. France: <https://www.statista.com/forecasts/1136348/facebook-users-in-france/>; worldwide: <https://www.statista.com/forecasts/1136351/facebook-users-in-the-world/>. 2018 modeled historical observations; country and worldwide series updated 2026-10-01 and 2026-10-01; accessed October 17, 2026.
- Statista (2026b). Instagram users in france and worldwide: Key market indicators. France: <https://www.statista.com/forecasts/1138801/instagram-users-in-france/>; worldwide: <https://www.statista.com/forecasts/1138856/instagram-users-in-the-world/>. 2018 modeled historical observations; country and worldwide series updated 2026-09-28 and 2026-09-28; accessed October 17, 2026.
- Statista (2026c). LinkedIn users in france and worldwide: Key market indicators. France: <https://www.statista.com/forecasts/1147058/linkedin-users-in-france/>; worldwide: <https://www.statista.com/forecasts/1147197/linkedin-users-in-the-world/>. 2018 modeled historical observations; country and worldwide series updated 2026-09-28 and 2026-09-29; accessed October 17, 2026.
- Tarduno, M. (2025). Elasticities and tax incidence in urban ridesharing markets: Evidence from chicago. *Journal of Urban Economics*, 150:103809.
- TripAdvisor (2020). Form 10-k for the year ended december 31, 2019. https://www.sec.gov/Archives/edgar/data/1526520/000156459020005269/trip-10k_20191231.htm. Management discussion and analysis.
- United States Department of the Treasury (2021). Joint statement from the united states, austria, france, italy, spain, and the united kingdom, regarding a compromise on a transitional approach to existing unilateral measures during the interim period before pillar 1 is in effect.
- United States Trade Representative (2019). Report on france’s digital services tax. Technical report, Office of the United States Trade Representative.
- United States Trade Representative (2021). Suspension of tariff action in france digital services tax investigation.

A Firm identification and website linkage

This appendix documents how I build the research roster of 35 corporate groups that are likely liable for the French DST and how I link those groups to websites in the WhoTracksMe (WTM) traffic universe. The work proceeds in two stages. The first stage starts from the U.S. Trade Representative’s list of 27 groups, assembles a broad universe of candidate and peer firms with financial data, and uses financial screens and a business-model review to select eight additional groups. The second stage assigns each WTM domain to its December 2018 owner, groups domains into website families, and applies coverage and activity screens to obtain the liable website families used in the traffic analysis. Table A.5 at the end of the appendix summarizes the populations that result.

A.1 Starting point: the USTR list

The starting point is the table of company groups “expected to be covered” by the French DST in the USTR’s Section 301 report ([United States Trade Representative, 2019](#)). The USTR drew that table from an economic-impact study commissioned from Deloitte/Taj and from the French National Assembly’s report on the bill. It names 27 groups: Airbnb, Alibaba, Alphabet, Amadeus, Amazon, Apple, Axel Springer, Booking Holdings, Criteo, eBay, Expedia, Facebook, Groupon, Match Group, Microsoft, Rakuten, Randstad, Recruit, Sabre, Schibsted, Snap, Travelport, Twitter, Uber, Verizon, Wish (ContextLogic), and Zalando. Because the USTR compiled the list to support its finding that the tax discriminates against U.S. firms, I treat it as a starting point that requires validation in two directions: groups on the list that may not meet the statutory criteria, and groups that meet them but are absent from the list.

A.2 Candidate universe and source databases

Candidate list. I first compiled a deliberately over-inclusive screening list of 130 groups from public sources: the USTR and Deloitte materials, the statutory text, parliamentary materials, and ministerial statements, with assistance from a large language model. The list contains all 27 USTR groups. The remaining rows are candidates to be confirmed or rejected against accounting data. Each row records the group’s in-scope service category (digital intermediation, or targeted advertising and the sale of user data), a confidence tier (25 high, 39 medium, 66 low), and a threshold band. Forty-three rows are expected to clear the statutory thresholds of €750 million worldwide and €25 million in France. The other 87 were added under a lowered safety band of €500 million and €5 million. The list also retains groups that were in scope on 1 January 2019 but were later acquired, renamed, merged, or closed, and it records the legal entity that filed for the relevant year. Over-inclusion is intentional. A false positive at this stage only adds records that the screens later remove, whereas a group omitted at this stage cannot be recovered.

Industry-peer universe. Each candidate is resolved to a database identifier: listed firms by ticker to a Compustat `gvkey`, and private firms by name and country to an Orbis BvD identifier. I then expand the

universe to every firm with roughly €500 million or more in global operating revenue whose primary industry code falls in the union of the NAICS and SIC codes carried by the resolved candidates. Firms come from Compustat, which covers listed firms worldwide, and from Orbis, which adds private firms and firms with a French presence (S&P Global Market Intelligence, 2025; Bureau van Dijk, a Moody's Analytics company, 2025). Firms whose business model appears out of scope, such as first-party retailers, pure software providers, and subscription streaming services, are tagged rather than dropped, so that the screens can separate them later. The resulting financial panel covers 3,895 firm identifiers over 2014–2025: 105 resolved primary candidates, 3,788 industry peers, and two duplicate French-shell identifiers. Compustat amounts are converted to euros using Orbis's own exchange rates, and all screens use fiscal year 2018.

French footprint. Statutory French revenue is not reported in consolidated accounts, so I reconstruct each group's 2018 French footprint from the Orbis ownership tree as of 2018 and the unconsolidated accounts of its French entities. The default aggregation sums operating leaves of the tree and excludes holding companies where possible; a French group that reports its own activity can enter through its own accounts. Four alternatives retain operating parents, sum all directly attributed French entities, select the largest French entity, or follow the global-ultimate-owner path. These are alternative accounting footprints used to assess attribution and double counting. None of them measures turnover apportioned by the statutory user-location rules.

A.3 What the financial screens establish

The screens combine company size, industry classification, business-segment composition, and the French footprints. I use two industry-code sets. The broad set (49 NAICS and 24 SIC codes) is the union of the candidates' codes. The strict set (12 NAICS codes) keeps only codes that correspond to digital intermediation or advertising services. It drops codes that the broad set inherits from conglomerate candidates, such as Apple's equipment-manufacturing code or Amazon's e-commerce retail codes. The broad screen requires worldwide revenue above €750 million in a broad in-scope industry. The French screens additionally require recorded French revenue above €25 million under one of the footprint aggregations. The segment-based screen applies both thresholds to revenue scaled by the share of segment sales in strict in-scope industries. Because the broad code set covers about 97 percent of the pre-curated universe, the global size threshold, rather than the industry criterion, is what binds in the broad screen.

Table A.1 reports the screens' overlap with the USTR list. Missing segment data, unresolved French entities, and recorded French sales below €25 million account for some gaps. Agreement with the USTR list describes overlap between classification methods, rather than the sensitivity or specificity of a verified liability test.

Several cases show why documentary review remains necessary. Airbnb, Alibaba, and eBay have recorded French sales below €25 million in the report despite appearing on the USTR list. Their local accounts need not capture the revenue attributable to French users. Rakuten's French entities appear in the

Table A.1: Financial-screen overlap with the 27 USTR groups

Screen in the archived FY2018 report	Groups flagged	USTR groups recovered
Broad size and industry screen	893	25
French operating-leaf screen	118	15
French operating-parent screen	129	16
All directly attributed French entities	131	16
Largest French entity	127	16
Segment-based, operating-parent screen	50	10

Notes: Source: the archived FY2018 liability-screen report underlying the institutional review. “Recovered” means flagged by the indicated screen and present on the USTR list; it does not mean independently verified as liable. The broad screen applies the global size and industry criteria. The French screens additionally use the indicated accounting footprint. The segment-based screen adds the report’s business-segment criterion. Global sales and French subsidiary sales can differ from covered revenue under the tax law. The global-ultimate-owner variant is excluded here because the report identifies attribution problems in that variant.

2018 ownership tree, but their accounts cannot be reconstructed from the unconsolidated Orbis records. Expedia’s French revenue is attributed to a sibling holding company, Liberty Expedia Holdings. The USTR list’s industry codes can also be misleading: Verizon is coded as a telecommunications firm, but it is on the list for the advertising business of Verizon Media (Yahoo and AOL). In the other direction, retailers, software companies, and payments providers can meet broad size and industry criteria while earning revenue from services outside the tax base. Screening on the global threshold and a broad in-scope code alone flags groups such as Walmart and Costco.

These considerations support retaining a documented research roster rather than applying an automatic liability indicator from financial accounts. I therefore keep all 27 USTR groups regardless of their screen results, because where a screen fails to confirm a USTR group the cause is missing French or segment data rather than evidence that the group is out of scope. The same considerations explain why parent revenue shares should be interpreted as measures of business composition. The shares are not estimates of the share of revenue on which a firm actually paid the French tax.

A.4 Selecting additions to the USTR roster

The financial screens identify candidates for business-model review. The final research roster retains the 27 USTR groups independently of their accounting-screen results and adds eight groups selected through two parallel screens. The 893 firms passing the broad size and industry screen therefore form a search pool for potential additions, rather than a roster that is reduced mechanically to 35 firms.

The first path uses worldwide revenue, recorded French revenue, and industry codes. Of the 893 firms passing the broad screen, 129 also pass the French revenue requirement using the operating-parent aggregation. Removing the 16 USTR groups and two duplicate representations of USTR groups—Match Group—OLD and Liberty Expedia Holdings—leaves 111 potential additions. Restricting these firms to the strict industry-code set removes 79 broad-code-only matches and leaves 32 candidates. Business-model review classifies eight as defensible additions, nine as borderline, and 15 as primarily first-party software, IT, or

payments businesses. The eight additions are Omnicom, Publicis, Interpublic, Dentsu, Cheil, TripAdvisor, ANGI, and Just Eat.

The second path uses business-segment revenue shares to narrow the activity criterion. The operating-parent segment screen flags 50 firms, including ten USTR groups. Removing those ten groups, the same two duplicate representations, and 17 firms that pass only the broader segment definition leaves 21 candidates passing the narrower segment criterion. Business-model review retains five: Omnicom, Publicis, Interpublic, TripAdvisor, and ANGI. Four candidates remain borderline and 12 primarily provide first-party software, IT, or payments services. The industry-based path recovers three additional groups—Dentsu, Cheil, and Just Eat—that lack the segment splits needed for the segment-based test. Their absence from that test reflects missing data rather than a finding that their activities fall outside the tax base.

The business-model review asks a single question of each surviving firm: does its revenue come predominantly from one of the taxed services—digital intermediation, targeted advertising based on user data, or the sale of user data—or from activities the tax excludes? I answer it from firms’ descriptions of their businesses and segments in annual reports and from the strict segment share. I sort firms into three tiers: defensible additions, borderline cases, and mechanical passes whose codes fall in scope but whose business does not. Table A.2 lists the outcome. The exclusions reflect the distinction between an industry or segment code and the service actually supplied. Firms such as IBM, Oracle, SAP, Autodesk, ADP, and payments providers can pass a code-based screen while primarily earning revenue from their own products or services. The review also leaves mixed or ambiguous businesses outside the additions: TUI, Thomas Cook, H.I.S., and Fosun Tourism predominantly supply first-party travel services; IHS Markit and FactSet sell financial rather than user data; GfK’s market-research data is not clearly user data generated on a digital interface; JCDecaux’s advertising is primarily out-of-home; Adobe combines advertising-related activities with first-party software; and Comcast’s in-scope advertising is a thin share of its revenue (a strict segment share of 0.12). These classifications explain why passing the accounting thresholds alone does not establish likely liability for every candidate.

Two of the eight additions, TripAdvisor and Just Eat, were already on the hand-compiled candidate list as medium-confidence intermediaries. ANGI and the five agency groups entered only through the industry-peer expansion. This is the main practical payoff of widening the universe beyond the hand-compiled candidates. Two further checks guard against missed groups. Taking the union of candidates across all five French aggregation rules adds firms whose French revenue comes from misattributed ownership-tree paths rather than from digital services. Among large firms that pass the size threshold but have no resolved French revenue, the missing French revenue is almost always a true zero for firms without French operations rather than a data gap.

The resulting roster contains the 27 USTR groups plus five additions supported by both screening paths and three recovered through the industry-based path, for 35 groups in total. More generally, the roster records likely rather than confirmed liability: the accounting and business-model evidence does not verify the statutory revenue base or individual tax payments. Two outside sources are consistent with a list of this

Table A.2: Business-model review of non-USTR firms passing the financial screens

Tier	Firms
<i>Defensible additions (8)</i>	
Intermediation	TripAdvisor (incl. Viator and TheFork), ANGI (HomeAdvisor), Just Eat
Advertising	Omnicom, Publicis, Interpublic, Dentsu, Cheil
<i>Borderline, not added</i>	
Travel, mostly first-party	TUI, Thomas Cook, H.I.S., Fosun Tourism
Data, not user data	IHS Markit, FactSet, GfK
Mixed or thin advertising	JCDecaux (out-of-home), Adobe, Comcast
<i>Code-based passes outside the taxed services</i>	
	IBM, Oracle, SAP, ADP, Fidelity National Information Services, Conduent, Open Text, Akamai, Corpay, Autodesk, Exela, WEX, and others

Notes: Firms are the union of the non-USTR survivors of the strict industry-code path (32 firms) and the strict segment-share path (21 firms) described in the text. The five advertising-agency groups fall in the same service class as Criteo on the USTR list; for agencies, only remuneration for covered targeted-advertising services enters the tax base, so total agency turnover is not a measure of taxable receipts. Tiers record a judgment about each firm's predominant business in 2018, not a verified statutory classification.

size. In answer to a parliamentary question, the French tax administration reported that 27 firms paid the DST for 2019, 28 for 2020, and 36 for 2021 ([Assemblée Nationale, 2023](#)). Its unit of count is the contributing legal entity, which need not map one-for-one to a corporate group, so I read the totals as indicating the right order of magnitude rather than as a firm-by-firm validation. Separately, a search of corporate filings and investor materials finds explicit references to French DST liabilities for Booking, ANGI, TripAdvisor, and Match (Appendix B).

A.5 Construction of the corporate revenue table

Table 2 combines two sources. Consolidated revenue and its split by activity (columns 1, 3, and 4) come from a structured extraction of each group's own FY2018 financial reports. French-entity turnover (column 2) comes from Orbis subsidiary accounts. Columns 5–8 are ratios of these four columns. Apart from a small number of documented adjustments, listed in Table A.3, every entry follows mechanically from the two sources.

Filing extraction. For each group, I collected the documents covering its FY2018 fiscal year:

- SEC filers: the annual report on Form 10-K or 20-F, the following year's report for restated comparatives, and quarterly earnings releases. Uber, Airbnb, and Wish were private in 2018, so their IPO prospectuses take the place of an annual report.
- The 12 groups that do not file with the SEC: annual reports, results releases, and investor presentations from their own websites.

In total, this amounts to 324 SEC documents and 52 home-country documents. FY2018 is the fiscal year

ending between June 2018 and May 2019. Most groups report on a calendar year, but Microsoft’s fiscal year ends in June 2018, Apple’s in September 2018, and Alibaba’s and Recruit’s in March 2019.

Figures were extracted with AI assistance following a written specification. Each figure is recorded exactly as printed, with its scale, currency, document locator, and a verbatim quote of the source row. The quotes are checked automatically against the document text. An item a firm does not disclose is recorded as a disclosure gap in the firm’s own words, never as a zero. Each group’s total revenue is checked against Compustat, and gaps above one percent are flagged. Two such gaps reflect entity differences rather than errors. Compustat’s FY2018 record under the project’s identifier for Match Group is its then parent, IAC, so the table uses the standalone Match Group’s filings. Cheil’s English-language releases label gross profit as revenue (discussed below). Figures are taken as first reported in FY2018 documents; restated comparatives from FY2019 reports are used only where no first-reported figure exists.

Revenue classification. Each extracted revenue line is assigned to one of six categories:

- digital advertising inventory sold by the firm;
- advertising-agency fee income;
- first-party online sales of the firm’s own inventory, recorded gross;
- intermediation revenue from marketplaces, travel, mobility, delivery, classifieds, jobs, and local services, recorded as net commissions and fees;
- fees of the travel distribution systems (Amadeus, Sabre, and Travelport);
- other revenue, such as hardware, software, cloud services, subscriptions, and traditional staffing.

Column 3 (“Ads”) is the sum of digital advertising and agency fees. Column 4 (“Intermediation”) is the sum of intermediation and distribution-system fees. First-party sales and other revenue enter neither column, so columns 3 and 4 need not sum to column 1.

Firms often print several alternative breakdowns of the same revenue—Expedia by service type and by business model, and Match Group by region and by brand—and summing across them double-counts. For each group I therefore use a single breakdown: the table whose lines sum most closely to total revenue. Where a firm’s tables mix subtotals with their components, as with Alibaba, the lines were selected by hand. For all 35 groups, the chosen breakdown sums to within two percent of total revenue.

Exchange rates. Amounts are converted to euros using the average of Orbis’s 2018 month-end exchange rates. The same calendar-2018 average applies to every group, including those whose fiscal year ends in March 2019. The cross-check against Compustat, by contrast, uses an average over each firm’s own fiscal year; the difference does not affect the revenue shares, which are ratios within a firm.

French-entity turnover. Column 2 sums FY2018 turnover from the unconsolidated Orbis accounts of the French entities attributed to each group in its 2018 ownership tree. The tree is held fixed at 2018, and a reported zero is treated as missing. French entities file calendar-year accounts, so for groups whose fiscal year does not end in December the numerator and denominator of column 8 cover periods that differ by up to a quarter. This turnover is not sales to French users, and it is not the DST base. For most platforms the French entity is a service company that bills its parent on a cost-plus basis, while revenue from French users is booked by an Irish, Luxembourg, or U.S. principal. Thus Alphabet’s, Amazon’s, and Booking’s French entities account for 0.4 percent or less of group revenue. Several entries need specific caveats:

- Amazon’s French retail activity runs through a French branch of a Luxembourg company and is absent from these accounts.
- Criteo and Publicis are headquartered in France, so their French entities include the group parent and column 8 reflects incorporation rather than French activity.
- Amadeus’s French turnover, about €7.4 million per French employee, reflects billing to the group by a research and service center rather than local sales.
- Randstad’s and Recruit’s French entities are staffing businesses whose turnover reflects placed temporary workers.

Five groups—Rakuten, ContextLogic, Zalando, ANGI, and TripAdvisor—have no French subsidiary accounts in Orbis and show dashes. Column 2 counts all directly attributed French entities. It is therefore related to, but not identical to, the alternative French footprints used in the liability screens of Table A.1. Column 8 divides column 2 by column 1.

Adjustments. Table A.3 lists the entries that depart from the mechanical construction, together with the reason for each.

A.6 From corporate groups to websites

Linking the roster to traffic requires a mapping from WTM domains to corporate owners. WTM reports traffic at the domain level and carries no ownership information. Its company field identifies the operators of trackers embedded in a page, not the owner of the page. The only bridge is therefore a domain-by-domain ownership classification. I construct it with AI-assisted labeling and researcher review in four steps. Throughout, a domain is assigned to the group that owned it in December 2018, and that assignment is held fixed over the whole sample period. Treatment is thus defined by pre-announcement ownership, and later transactions do not reassign domains.

Table A.3: Adjusted entries in the corporate revenue table

Group	Adjustment and reason
Verizon	Advertising is Oath (Verizon Media) revenue of \$7.7 billion, which the 10-K reports in its text rather than in a revenue table; the remainder of Verizon’s revenue is telecommunications.
Snap	Advertising is approximated as 99 percent of total revenue, the share Snap reports; its filings give no separate advertising amount.
Cheil	Total revenue is taken from Orbis. Cheil’s English-language releases report gross profit under the label “revenue”, and its gross sales appear only in the Korean audited report. Its advertising amount is omitted because the available figure is on the gross-profit basis.
Apple, Randstad	No split is shown. Apple does not report App Store revenue separately from its other services, and Randstad does not separate digital from traditional staffing revenue.
Rakuten	Advertising revenue is embedded in mixed revenue lines and is not separately allocated, so the advertising entry is zero by construction rather than by disclosure.
Match Group	Dating subscriptions are classified as other revenue, so intermediation is zero. Subscription revenue is likely within the tax base; taking it into account would place Match’s in-scope revenue above the global threshold.
Agencies	Agency fee income includes traditional as well as digital advertising services, so column 3 overstates the agencies’ digital advertising revenue. Only remuneration for covered targeted-advertising services enters the tax base.

Step 1: domain labeling. The initial pass labels the 2,794 distinct domains in the December 2018 WTM release. Each domain receives an ownership label (whether a roster group owned it in December 2018, and which one) and three activity flags following the statutory definitions: digital intermediation, targeted advertising, and the sale of user data. The labeling was carried out by language-model agents, each assigned a chunk of about 50 domains. All agents received the same written instructions and the same reference document. The reference document sets out the statutory scope rules with in-scope and out-of-scope examples. It also contains a catalog of the 27 USTR groups plus TripAdvisor, which was then under review as a borderline addition. For each group, the catalog lists its major operating domains and brands as of December 2018. Agents were instructed to recognize country-code variants and subsidiary brands not listed in the catalog and to use web searches for unfamiliar domains. They were told to apply the definitions to what each domain did in December 2018 rather than today, and never to assign a parent they could not confirm. An unconfirmed ownership was to be labeled non-labile with a note. The reference document flags dated ownership changes near the treatment date, including Flickr’s sale by Verizon in April 2018, Recruit’s acquisition of Glassdoor in June 2018, PriceMinister’s migration to Rakuten.fr in June 2018, and Schibsted’s ownership of its classifieds sites before the 2019 Adevinta spin-off. A script checked the combined output for complete coverage, valid values, and internal consistency, and 45 borderline rows were set aside for review.

The labeling was then extended to every domain that ranked among the top 2,000 in France, Germany, the European aggregate, the United States, or the global aggregate in any month before March 2022. This extended universe has 6,841 domains in 6,262 first-label families. The 3,694 families not covered by the first pass were represented by their most popular domain and labeled in two tiers. In a triage tier, agents labeled domains from prior knowledge without web search, in chunks of 150, and recorded their confidence in each

ownership call. A domain went to a second, research tier with web search if the triage agent suspected or could not rule out ownership by a roster group, or if a roster-owned domain's activity was uncertain. Twenty-two domains reached the research tier. A deterministic match against the catalog's domains and country-code variants served as a backstop for missed ownership. Labels were then propagated from each family's representative to the family's other domains, and a stratified sample of 42 rows was checked by hand. The result assigns 363 domains in 219 families to 26 of the 28 catalogued groups. Amadeus and Travelport, both business-to-business distribution systems, own no domain in the extended universe.

Step 2: firm crosswalk and reconciliation with the 35-group roster. The labeled domains are joined to firm identifiers in the financial panel through a company-to-identifier bridge. At this step I revisited ownership edges sensitive to mergers and acquisitions and recorded the basis, confidence, and notes for each edge. Applying the December 2018 rule corrected a legacy mislabel: Avito and OLX were owned by Naspers rather than Schibsted in 2018, and they are not treated. It also keeps Leboncoin and the other classifieds sites with Schibsted rather than with Adevinta, maps Verizon Media's domains to Verizon Communications, and assigns Glassdoor to Recruit. TripAdvisor's domains and Match Group's dating brands (including LoveScout24 and Twoo) were confirmed by the researcher. The crosswalk was then reconciled against the final 35-group roster. Every group in the crosswalk is on the roster. Seven roster groups operate no consumer-facing website: the five advertising agencies, Amadeus, and Travelport. Their exposure is financial only. The labeling catalog predated two additions with consumer websites, ANGI and Just Eat, and their domains were added to the crosswalk by hand: HomeAdvisor is assigned to ANGI, and Just Eat's own domains to the standalone Just Eat plc that existed in 2018. The Takeaway.com brands (Lieferando, Thuisbezorgd, Pyszne) and Grubhub were separate firms in 2018 and are not assigned to Just Eat. The resulting crosswalk links 221 families to 28 groups.

Step 3: committee classification of the traffic universe. The traffic analysis uses a narrower universe of domains with adequate coverage, which was classified again in more detail before any estimates were produced. This universe contains the 1,854 domains observed in both the French and global WTM panels in at least three months of April–November 2018 and at least six months of the post-announcement window. It adds every domain labeled as roster-owned that is observed at all in the window, for 2,020 domains in total. Each domain was classified independently by three language-model agents with distinct roles. The first classified the product and the domain's role, distinguishing consumer destinations from infrastructure, static-content, authentication, developer, and enterprise domains. The second classified the revenue model and scored exposure to each taxed activity. The third assessed competitive closeness to the roster groups' services. Agents saw only pre-treatment or time-invariant context. Batches kept all domains of a family together so that family membership could be curated consistently. A separate adjudicating model merged the three classifications field by field. It accepted unanimous and majority labels, resolved material disagreements on the evidence, and flagged low-confidence rows for review. An auditing model then re-examined

766 rows in three tiers: every proposed low-exposure control, every row with an ownership or family-collision flag, and every row flagged for review. The 24 rows that remained unresolved after the audit were decided by the researcher. Those rulings adopted four standing rules:

- A joint-venture interest carries treatment where ownership is established.
- A domain whose ownership changed at the treatment date for reasons predating the DST stays treated but leaves the primary sample.
- Domains with unverifiable or unaffiliated operators are excluded.
- Unresolvable ownership among exclusively untreated candidates is immaterial.

Treatment status itself comes from the step-1 ownership labels and changes only through these rulings. For example, the committee flagged Tradera as an eBay site that the source inventory had missed; a later review found that it was not owned by eBay in 2018, and it is dropped from the liable sample below. The Curse network sites that Amazon sold to Fandom in December 2018 were kept treated but excluded from the primary sample. The classifications were then frozen, and every later change is recorded in a changelog that notes whether it was made before or after estimates were seen. The frozen metadata contains 362 treated domains in 220 families across the 26 groups.

Step 4: website families. A website family pools a brand’s domains under its first domain label, so that `google.fr`, `google.de`, and `google.com` form one family. Pooling neutralizes traffic moves between country domains, such as Google’s consolidation of country-code traffic at the end of 2018. It also handles brands whose French storefront is a separate domain. Membership is curated rather than mechanical: a domain joins a family only if its owner matches the family’s, which keeps, for example, Naspers’s `olx` domains out of a family containing Schibsted’s `olx.com.br`. Family traffic in a month is the sum of popularity over the member domains observed that month, computed before taking logs. One recorded rebrand is pooled explicitly: Verizon’s `huffpost.com` is merged into the `huffingtonpost` family, because unpooled the rebrand appears as a large spurious increase in traffic.

A.7 From linked websites to the French liable sample

Table A.4 traces the treated families from the frozen metadata to the French sample. The first restrictions remove families that cannot carry a consumer traffic outcome. Two treated domains sit outside the curated family roster. Thirty-four families consist only of domains classified as infrastructure or non-consumer services, such as content-delivery, authentication, developer, and enterprise domains. Ten all-non-consumer families with substantial French traffic before the announcement are readmitted to the later hand review, so that their treatment is recorded as an explicit decision rather than an upstream filter: GitHub, Visual Studio, Microsoft Online, SharePoint, Googleusercontent, Android, Oath, Gamepedia, Hearthpwn, and Futhead.

Fifty-two further families have no French traffic in April–November 2018. Tradera is then dropped because it was not owned by eBay in 2018.

Table A.4: From roster-linked website families to the French liable sample

Step	Families	Groups
Treated families in the frozen domain metadata	220	26
On the curated family roster	218	26
At least one consumer-facing member domain (incl. 10 readmitted for review)	184	23
Some French traffic in April–November 2018	132	22
Hand-coded website classification, after dropping Tradera	131	22
Presence screen: at least four valid months in April–November 2018	71	21
Retained after activity review	50	20
French sample, restoring five French-oriented sites	55	20

Notes: Group counts exclude one treated domain with an unrecorded parent. The presence screen counts months in which both French and constructed rest-of-world traffic are usable. The five restored sites—La Centrale, Leboncoin, Logic-Immo, SeLoger, and Meetic—fail that screen only because the constructed rest-of-world measure is nonpositive for these predominantly French sites; four have all six June–November 2018 months of French traffic and Meetic has five. Steps after the presence screen involve recorded judgments rather than uniform rules, as described in the text.

Website classification. The 122 remaining websites with a classification row were hand-coded as of 2018, using AI-assisted drafts confirmed by the researcher site by site. Sixty-two of them had been coded in an earlier review. Each website is coded for whether it materially carried targeted advertising and whether it materially monetized paid digital intermediation, together with its primary traffic driver (shopping or content), its primary supply model (first or third party), and a sample decision. The coding uses a written codebook:

- *Targeted advertising:* paid third-party placements selected using user information, such as interest, behavioral, location, or search-query targeting, or sponsored listings.
- *Paid digital intermediation:* monetized user-to-user connections, such as commissions, listing or access fees, and paid matching features.
- *Materiality:* ancillary features are coded as absent.

Each row records its caveats and selected sources, such as annual reports, archived snapshots of the site, and media kits. The labels are qualitative judgments about a site’s dominant activity, not measured revenue shares.

Presence screen and activity review. The presence screen keeps the 71 families with at least four months in April–November 2018 in which both French and rest-of-world traffic are usable. A review of these 71 sites rates each for targeted advertising and paid intermediation as none, weak, or material. A site is retained if either activity is material. This review agrees with the hand-coded sample decisions for 65

sites. It retains six that the hand coding had excluded: Futhead, Gamepedia, and Hearthpwn, which are ad-supported Curse network sites, and OkCupid, Plenty of Fish, and Twoo, which are dating sites combining advertising with paid matching. The 21 excluded sites are primarily first-party products, communications services, or infrastructure: Apple, iCloud, Microsoft’s corporate site, Office, Office 365, Skype, Xbox, Minecraft, Messenger, WhatsApp, Periscope, Verizon Wireless, Blogger, Prime Video, and the readmitted infrastructure domains other than the three Curse sites. The resulting 50 families belong to 20 groups.

Restored French-oriented sites. Five French-oriented sites fail the presence screen only because the constructed rest-of-world measure is nonpositive for sites whose traffic is almost entirely French. I restore them, because the French-traffic outcome does not use that measure: La Centrale, Leboncoin, Logic-Immo, SeLoger, and Meetic. PriceMinister, the sixth site failing for this reason, is not restored: its domain migrated to Rakuten.fr in June 2018, so its pre-announcement series reflects the rebrand. The French sample therefore contains 55 families under 20 groups. Table 3 lists them, along with additional roster websites that pass the activity review but have only one to three positive French observations in June–November 2018. Families generally have at least four observed months in June–November 2018; Futhead is retained with three, covering September–November. The roster is fixed by these recorded decisions rather than generated by a uniform minimum-history rule. Available January–December 2019 observations enter the traffic panel without a complete post-announcement coverage requirement. Annual estimates compare September–November 2018 with January–December 2019, excluding December 2018.

Roster groups without sample websites. Fifteen of the 35 groups contribute no family to the French sample, for four reasons:

- *No consumer-facing website (7):* the five agencies, Amadeus, and Travelport.
- *Sites first observed after treatment (2):* Just Eat’s and ANGI’s sites first appear in French WTM data in 2019 (`just-eat.co.uk` in February, `just-eat.fr` in July, and `homeadvisor.com` in September), so they have no pre-announcement traffic.
- *Treated domains classified as non-consumer (3):* Criteo, Snap, and Sabre.
- *Insufficient French pre-announcement coverage or first-party activity (3):* Randstad’s sites (Monster and Military.com) have no French traffic in April–November 2018. Groupon’s sites fail the presence screen. Apple’s domains are first-party hardware retail and storage services and are excluded at the activity review.

Absence from the regression sample should not be interpreted as evidence of non-liability.

Nonliable comparison families. Comparison families are built from the same labeled universe. They are domains labeled as not owned by a roster group, excluding Wikipedia, which serves as the normalizing

site. These domains are pooled into families by first label in the same way. A comparison domain whose first label coincides with a liable family keeps its full domain as its key, so that, for example, Yahoo Japan stays separate from Verizon’s Yahoo. The 638 comparison families are those observed in all six months of June–November 2018. They satisfy the empirical coverage requirements and are distinguished by observed advertising exposure; the design does not treat all other websites in the underlying traffic universe as unexposed controls.

A.8 Limitations of the linkage

This procedure produces a research classification rather than an independent verification of statutory liability. Three limitations follow from how it was built.

- *Non-liable labels.* The labeling instructions assign an unconfirmed or conflicting ownership case to the non-liable category with an uncertainty note. A non-liable label therefore need not establish that ownership by a roster group has been affirmatively ruled out.
- *The eight additions.* The labeling catalog covered the 27 USTR groups and TripAdvisor. Domains of the other seven additions were not searched for systematically, and ANGI’s and Just Eat’s domains were added by hand. Because the five agencies operate no consumer destinations and ANGI’s and Just Eat’s sites lack pre-announcement French traffic, I do not expect this to affect the French sample. It could, however, place a domain owned by one of these groups among the candidate comparison domains.
- *Documentation.* The saved records contain the labeling instructions and reference catalog, assignment notes, the classification and audit outputs, the researcher’s rulings, and selected supporting sources. They do not provide a complete archived ownership citation for every domain.

The main advertising-composition analysis retains all 55 liable families and measures parent revenue composition using fiscal-year 2018 accounts. The 35-group institutional roster and the 28-firm disclosure inventory in Appendix B serve different purposes from the traffic sample, as Table A.5 summarizes.

Table A.5: Distinct populations used in the institutional and empirical work

Population	Firms or parents	Website families
Institutional research roster	35	–
Disclosure inventory, including home-country firms	28	–
SEC-filer component of disclosure inventory	21	–
Roster groups linked to websites in the WTM crosswalk	28	221
French liable traffic sample	20	55
Main advertising-composition sample	20	55

Notes: These are populations with different coverage requirements, not a sequence of nested sample restrictions. Disclosure counts refer to firms represented in the inventory, not to firms with an available filing in every year. The crosswalk row counts first-label website families in the extended labeled universe before any coverage screen. The final two rows describe the same liable-family roster used in the headline and main advertising-composition analyses. No aggregate traffic outcome is imputed to groups lacking observed websites.

B Corporate disclosures and tax collections

B.1 Sources and coverage

The disclosure inventory contains 28 firms, including 21 SEC filers and seven firms covered through home-country reports. It records fiscal years 2019–2024, with the initial review also consulting 2018 filings. SEC materials include annual and quarterly reports and earnings releases; the narrative search covers risk factors, management discussion, legal proceedings, and financial-statement notes. For each firm-year, the inventory records whether a digital services tax is mentioned, whether France is identified, whether a numerical amount is supplied, and what that amount measures.

Reported amounts include expenses, changes in broader expense categories, payments, and contingent exposures. I record their accounting meaning and geographic coverage separately. An amount covering several countries can be assigned to France only when the report provides a country breakdown.

Coverage varies with listings, delistings, ownership changes, and the availability and parsing of reports. A missing France-specific amount means that the review did not recover one; it provides no evidence of zero liability or payment.

B.2 DST expense disclosures

Table [B.1](#) presents four 2019 disclosures recorded in the review. Booking identifies \$36 million associated with French digital service taxes in its discussion of general and administrative expenses. ANGI identifies \$0.9 million in its European expense discussion. TripAdvisor records a \$3 million estimate immediately after describing the French law. The Match segment’s \$4.7 million figure, reported in IAC’s consolidated filing, is an increase in non-income taxes primarily related to the French tax; it is not a separately identified French tax total ([Booking Holdings, 2020](#); [ANGI Homeservices, 2020](#); [TripAdvisor, 2020](#); [IAC/InterActiveCorp, 2020](#)).

Table B.1: Selected 2019 disclosures and their accounting meaning

Firm or segment	Reported amount	Interpretation
Booking Holdings	\$36m	French digital service taxes identified in the general and administrative expense discussion.
ANGI	\$0.9m	Amount associated with the French tax in the increase in European general and administrative expenses.
TripAdvisor	\$3m	Estimated digital service tax expense disclosed in the discussion of France's 2019 law.
Match segment (IAC)	\$4.7m	Increase in non-income taxes primarily related to the French tax; the entire amount is not isolated as French DST.

Notes: Amounts are in the reporting currency. They are expense disclosures, not verified cash payments. Source: FY2019 Forms 10-K, management discussion, for [Booking Holdings \(2020\)](#); [ANGI Homeservices \(2020\)](#); [TripAdvisor \(2020\)](#); [IAC/InterActiveCorp \(2020\)](#). The Match observation is located in IAC's discussion of its MTCH segment.

Adevinta's 2022 annual report provides a country-specific example outside the SEC inventory. It describes payments on account of €11 million for 2022 and €22 million for 2019–2021 following a contested French ruling. The company appealed and expected to pay an additional €6 million in early 2023 ([Adevinta, 2023](#)). Its €33 million payment during 2022 thus covered several tax years while the dispute remained unresolved.

The four 2019 disclosures appear in discussions of operating expenses, where a search confined to the corporate income-tax footnote would miss them. They document costs associated with the tax but reveal little about changes in spending on particular services or geographic markets.

B.3 Government receipts

Government receipts show the scale of collections. A parliamentary response reports €277 million from 27 contributing enterprises in 2019, €375 million from 28 in 2020, and €474 million from 36 in 2021 ([Assemblée Nationale, 2023](#)). Table B.2 reproduces these totals. The response classifies nationality by the contributing enterprise's registered office, so its enterprise counts cannot be matched directly to the ultimate parents in the firm list.

Table B.2: French tax receipts and contributing enterprises, 2019–2021

Calendar year	Receipts (€m)	Contributing enterprises
2019	277	27
2020	375	28
2021	474	36

Notes: Source: DGFIP data reported in the government’s February 14, 2023 response to parliamentary question 2974 ([Assemblée Nationale, 2023](#)). Receipts concern budget line 1430. The administrative enterprise count is not assumed to equal a count of ultimate-parent groups.

Collections and reported corporate expenses cover different firms, periods, and accounting concepts. Comparing their totals therefore cannot establish how much of the tax appears in public filings.

C Advertising demand and tax responses

This appendix derives advertising demand and the tax responses for the content-consumption benchmark in Section 3.1.

C.1 Downstream demand for advertising

This section provides a derivation for downward-sloping demand. The assumptions are explicit: advertising helps firms reach customers, impression requirements increase with sales, and underlying product valuations remain unchanged. This approach emphasizes the idea that what social media platforms are doing is supplying an input into production of a final good.

Consider a competitive downstream product market. Let q denote product sales, $P_B(q)$ inverse consumer demand, and $P_S(q)$ marginal supply cost excluding advertising, with $P'_B(q) < 0$ and $P'_S(q) > 0$. Advertising is a required selling input that allows firms to reach customers. It does not change consumers' underlying product valuations.

Supporting sales q requires $A = \kappa(q)$ advertising impressions, where

$$\kappa(0) = 0, \quad \kappa'(q) > 0, \quad \kappa''(q) \geq 0.$$

The linear requirement $\kappa(q) = \bar{\kappa}q$, with $\bar{\kappa} > 0$, is a special case. At price p_A per impression, total downstream supply cost is

$$\int_0^q P_S(x) dx + p_A \kappa(q).$$

Product-market equilibrium therefore satisfies

$$P_B(q) = P_S(q) + p_A \kappa'(q).$$

Advertising enters the marginal cost of supplying the product. Since $\kappa'(q) > 0$, the requirement can be inverted to obtain $q = \kappa^{-1}(A)$. Inverse advertising demand is then

$$P_A(A) = \frac{P_B(\kappa^{-1}(A)) - P_S(\kappa^{-1}(A))}{\kappa'(\kappa^{-1}(A))}.$$

Writing $q = \kappa^{-1}(A)$, differentiation gives

$$P'_A(A) = \frac{[P'_B(q) - P'_S(q)] \kappa'(q) - [P_B(q) - P_S(q)] \kappa''(q)}{[\kappa'(q)]^3} < 0$$

on the positive-price domain $P_B(q) > P_S(q)$. Advertising expenditure is determined by this market equilibrium, rather than by an exogenous advertising budget. Assume P_B and P_S are twice continuously differentiable and κ is three times continuously differentiable, so P_A is twice continuously differentiable.

C.2 Conditional expenditure and impression volume

Assume that $\mathcal{L}$ and F are twice continuously differentiable and that the conditional minima defining $U(A)$ exist and vary smoothly on the relevant positive interior branch. For a fixed volume A , the conditional expenditure problem has Lagrangian

$$\mathcal{J} = g_U + \lambda [A - a\mathcal{L}(a, g_U)],$$

where λ is the multiplier on the impression requirement. Its first-order conditions imply

$$\lambda = \frac{1}{a\mathcal{L}_{g_U}} > 0, \quad \mathcal{L} + a\mathcal{L}_a = 0.$$

The envelope theorem gives $U'(A) = \lambda$. Hence total marginal expenditure is positive:

$$C'_A(A) = (F^{-1})'(A) + \frac{1}{a\mathcal{L}_{g_U}} > 0.$$

Holding A fixed, the constraint locally defines the required g_U as a function of a . At its stationary point, the second derivative of this required expenditure is

$$\left. \frac{\partial^2 g_U}{\partial a^2} \right|_A = -\frac{2\mathcal{L}_a + a\mathcal{L}_{aa}}{a\mathcal{L}_{g_U}} > 0.$$

Thus the strict conditional second-order condition requires $2\mathcal{L}_a + a\mathcal{L}_{aa} < 0$.

For reduced-form profit $\Pi(A; \tau) = (1 - \tau)P_A(A)A - C_A(A)$, the first-order and strict second-order conditions are

$$(1 - \tau) [P_A(A) + AP'_A(A)] = C'_A(A),$$

and

$$\Pi_{AA}(A; \tau) = (1 - \tau) [2P'_A(A) + AP''_A(A)] - C''_A(A) < 0.$$

Implicit differentiation yields

$$\frac{dA}{d\tau} = \frac{P_A(A) + AP'_A(A)}{(1 - \tau) [2P'_A(A) + AP''_A(A)] - C''_A(A)} < 0.$$

The numerator equals $C'_A(A)/(1 - \tau) > 0$. The expenditure and price responses follow from the production constraint, conditional expenditure function, and advertising demand:

$$\begin{aligned}\frac{dg_A}{d\tau} &= (F^{-1})'(A) \frac{dA}{d\tau} < 0, \\ \frac{dg_U}{d\tau} &= \frac{1}{a\mathcal{L}_{g_U}} \frac{dA}{d\tau} < 0, \\ \frac{dp_A}{d\tau} &= P'_A(A) \frac{dA}{d\tau} > 0.\end{aligned}$$

C.3 Advertising intensity and traffic

The possibility of a traffic increase depends on how advertising intensity and user-facing expenditure interact in the traffic function. In the multiplicatively separable case $\mathcal{L}(a, g_U) = H(g_U)V(a)$, optimal intensity is locally unchanged by the tax and traffic falls. A positive traffic response therefore requires a departure from this separable benchmark, with intensity falling sufficiently to offset the loss from lower user-facing expenditure.

Differentiating the conditional intensity condition gives

$$(2\mathcal{L}_a + a\mathcal{L}_{aa}) \frac{da}{d\tau} + (\mathcal{L}_{g_U} + a\mathcal{L}_{ag_U}) \frac{dg_U}{d\tau} = 0.$$

Since $dg_U/d\tau < 0$, intensity falls, rises, or is locally unchanged according to whether $\mathcal{L}_{g_U} + a\mathcal{L}_{ag_U}$ is positive, negative, or zero. The interior condition also implies $\mathcal{L}_a = -\mathcal{L}/a < 0$; this benchmark does not cover an interior optimum at which advertising increases traffic.

Substituting the intensity and expenditure responses into $dL/d\tau = \mathcal{L}_a(da/d\tau) + \mathcal{L}_{g_U}(dg_U/d\tau)$ yields

$$\frac{dL}{d\tau} = \frac{\mathcal{L}_{g_U}(\mathcal{L}_a + a\mathcal{L}_{aa}) - a\mathcal{L}_a\mathcal{L}_{ag_U}}{a\mathcal{L}_{g_U}(2\mathcal{L}_a + a\mathcal{L}_{aa})} \frac{dA}{d\tau}.$$

The denominator and $dA/d\tau$ are both negative. Using $\mathcal{L}_a = -\mathcal{L}/a$, traffic therefore rises precisely when

$$\frac{a^2\mathcal{L}_{aa}}{\mathcal{L}} + \frac{a\mathcal{L}_{ag_U}}{\mathcal{L}_{g_U}} > 1.$$

Traffic falls when the left-hand side is below one and is locally unchanged when it equals one. Equivalently, $L = A/a$ implies

$$\frac{d \log L}{d\tau} = \frac{d \log A}{d\tau} - \frac{d \log a}{d\tau}.$$

Since impressions fall, traffic can rise only if advertising intensity falls proportionately more than impressions.

For multiplicatively separable traffic, $\mathcal{L}(a, g_U) = H(g_U)V(a)$, with $H, V > 0$ on the relevant domain,

the intensity condition becomes

$$V(a) + aV'(a) = 0.$$

The strict conditional second-order condition selects a locally isolated solution independent of g_U . Thus $da/d\tau = 0$, and lower user-facing expenditure reduces traffic. All comparative statics in this appendix are evaluated at the maintained interior optimum.

D Intermediation costs, fees, and traffic

This appendix derives the tax responses for the transaction-intermediation model in Section 3.2. Intermediation expenditure is the operating cost of facilitating transactions. It does not independently shift buyer demand, seller supply, or service quality.

D.1 Transaction volume and operating expenditure

Assume that inverse buyer demand D^{-1} , inverse seller supply S^{-1} , and the transaction technology F are twice continuously differentiable on the relevant domains. The fee schedule and required expenditure are

$$P_Q(Q) = D^{-1}(Q) - S^{-1}(Q), \quad C_Q(Q) = F^{-1}(Q).$$

The demand and supply slopes imply $P'_Q(Q) < 0$. Since $F' > 0$ and $F'' \leq 0$, required expenditure has positive marginal cost and is weakly convex:

$$C'_Q(Q) = \frac{1}{F'(g_Q)} > 0, \quad C''_Q(Q) = -\frac{F''(g_Q)}{[F'(g_Q)]^3} \geq 0, \quad g_Q = F^{-1}(Q).$$

The platform chooses feasible transaction volumes with nonnegative fees. At a positive interior optimum of equation (7), the first-order condition is

$$(1 - \tau) [P_Q(Q) + QP'_Q(Q)] = C'_Q(Q).$$

The strict second-order condition requires

$$\Pi_{QQ}(Q; \tau) = (1 - \tau) [2P'_Q(Q) + QP''_Q(Q)] - C''_Q(Q) < 0.$$

Differentiating the first-order condition gives

$$\frac{dQ}{d\tau} = \frac{P_Q(Q) + QP'_Q(Q)}{(1 - \tau) [2P'_Q(Q) + QP''_Q(Q)] - C''_Q(Q)} < 0.$$

The numerator equals $C'_Q(Q)/(1 - \tau) > 0$, and the denominator is negative by the second-order condition. The reduction in transaction volume lowers required operating expenditure:

$$\frac{dg_Q}{d\tau} = C'_Q(Q) \frac{dQ}{d\tau} < 0.$$

D.2 Fees, prices, and traffic

Along the fee schedule, buyer demand, and seller supply, the price responses are

$$\begin{aligned}\frac{dp_Q}{d\tau} &= P'_Q(Q) \frac{dQ}{d\tau} > 0, \\ \frac{dp_B}{d\tau} &= (D^{-1})'(Q) \frac{dQ}{d\tau} > 0, \\ \frac{dp_S}{d\tau} &= (S^{-1})'(Q) \frac{dQ}{d\tau} < 0.\end{aligned}$$

The buyer-price increase and seller-compensation reduction sum to the fee increase: $dp_Q/d\tau = dp_B/d\tau - dp_S/d\tau$. Market clearing gives $S(p_S) = Q$, so the change in supply is

$$\frac{dS(p_S)}{d\tau} = S'(p_S) \frac{dp_S}{d\tau} = \frac{dQ}{d\tau} < 0.$$

Seller availability is distinct from aggregate supply and completed transactions. As in the main text, let $N = H(p_S)$ denote the number of active sellers, where H is continuously differentiable and $H'(p_S) > 0$. This participation assumption is additional to upward-sloping aggregate supply. Lower compensation then reduces seller participation:

$$\frac{dN}{d\tau} = H'(p_S) \frac{dp_S}{d\tau} = H'(p_S)(S^{-1})'(Q) \frac{dQ}{d\tau} < 0.$$

For a differentiable traffic function $L = \tilde{\mathcal{L}}(p_B, N)$, the equilibrium traffic response is

$$\begin{aligned}\frac{dL}{d\tau} &= \tilde{\mathcal{L}}_{p_B} \frac{dp_B}{d\tau} + \tilde{\mathcal{L}}_N \frac{dN}{d\tau} \\ &= \left[\tilde{\mathcal{L}}_{p_B} (D^{-1})'(Q) + \tilde{\mathcal{L}}_N H'(p_S) (S^{-1})'(Q) \right] \frac{dQ}{d\tau}.\end{aligned}$$

Because $dQ/d\tau < 0$, traffic falls precisely when the bracketed expression is positive. The maintained assumptions $\tilde{\mathcal{L}}_{p_B} < 0$ and $\tilde{\mathcal{L}}_N > 0$, together with downward-sloping buyer demand, upward-sloping seller supply, and $H' > 0$, are sufficient for this result. Higher buyer prices discourage visits, while fewer active sellers reduce the availability of offers that attracts users. If the bracket is negative, traffic rises despite fewer completed transactions; if it is zero, traffic is locally unchanged. All responses are evaluated at the maintained interior optimum.

E Website-level traffic effects

Table E.1 reports annual traffic-effect estimates for each of the 55 liable website families belonging to 20 corporate parents. For each family, it presents a raw synthetic-control contrast, the composition model’s prediction, and an empirical Bayes estimate combining the two. All three use the main French-only synthetic-control specification, with donors drawn from the 159 nonliable families without observed pre-announcement advertising connections to liable platforms.

Let $g_{it} = y_{it} - \hat{y}_{it}^0$ denote family i ’s observed-minus-synthetic log traffic gap. The raw annual contrast is

$$\hat{\delta}_i = \frac{1}{12} \sum_{t \in Q} g_{it} - \frac{1}{3} \sum_{t \in P} g_{it}, \quad P = \{\text{Sep.} - \text{Nov. 2018}\}, \quad Q = \{\text{Jan.} - \text{Dec. 2019}\}.$$

December 2018 is excluded. For each point estimate, synthetic weights are fitted using available June–November 2018 French traffic and applied to the post-announcement path. The full-pipeline inference procedure refits those weights in every bootstrap draw.

The model-implied effect is

$$m_i = \hat{\beta} + \hat{\theta}_C \frac{C_{j(i)} - \bar{C}}{s_C},$$

where $C_{j(i)}$ is advertising’s share of advertising plus intermediation revenue for family i ’s parent. The mean $\bar{C}$ and standard deviation s_C are calculated across the 55 liable families. The coefficients come from the original stacked composition regression. Families with the same composition therefore share the same model prediction, while their raw and EB estimates can differ.

The EB estimate is

$$\tilde{\delta}_i = r_i \hat{\delta}_i + (1 - r_i) m_i, \quad r_i = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{s}_i^2}.$$

I estimate the raw contrast’s sampling variance $\hat{s}_i^2$ using a website-specific baseline and constant 2019 shift, with a common AR(1) residual correlation and website-specific scales. Holding the original-model predictions fixed, I estimate the residual signal variance $\hat{\tau}^2$ by marginal likelihood. The reliability weight r_i determines how strongly the raw contrast is pulled toward its composition-model prediction. It is distinct from the synthetic donor weights: larger r_i gives greater weight to the raw website contrast.

The table reports $100\hat{\delta}_i$, $100m_i$ and $100\tilde{\delta}_i$ in log percentage points, and $100r_i$ in the final column. It groups families by their parent’s composition: pure advertising ($C = 1$; 13 families), mixed ($0 < C < 1$; 37 families), and pure intermediation ($C = 0$; five families). These are annual point estimates, calculated directly from the pre/post contrast. They need not equal averages of the separately fitted monthly EB estimates.

Table E.1: Annual website-level raw, model-implied and empirical
Bayes traffic effects

Website family	Corporate parent	Ad share (%)	Raw synthetic	Model-implied	Empirical Bayes	Weight on raw (%)
<i>Pure advertising (13 families)</i>						
google	Alphabet (Google)	100.0	+0.48	+16.48	+3.91	78.6
youtube	Alphabet (Google)	100.0	+17.31	+16.48	+17.19	85.2
meetic	Match Group	100.0	+23.85	+16.48	+20.54	55.2
okcupid	Match Group	100.0	+88.18	+16.48	+32.44	22.3
pof	Match Group	100.0	+2.54	+16.48	+15.44	7.4
twoo	Match Group	100.0	+38.37	+16.48	+18.79	10.6
facebook	Meta (Facebook)	100.0	-2.60	+16.48	-0.12	87.0
instagram	Meta (Facebook)	100.0	+41.19	+16.48	+32.01	62.9
twitter	Twitter	100.0	+40.66	+16.48	+32.26	65.3
aol	Verizon (Verizon Media)	100.0	-3.78	+16.48	+4.83	57.5
huffingtonpost	Verizon (Verizon Media)	100.0	+20.41	+16.48	+19.26	70.6
tumblr	Verizon (Verizon Media)	100.0	+20.80	+16.48	+18.78	53.2
yahoo	Verizon (Verizon Media)	100.0	+1.89	+16.48	+4.37	83.0
<i>Mixed (37 families)</i>						
alibaba	Alibaba	59.1	+15.93	+2.57	+6.19	27.1
aliexpress	Alibaba	59.1	+12.24	+2.57	+5.26	27.8
taobao	Alibaba	59.1	-15.50	+2.57	+0.49	11.5
amazon	Amazon	19.1	+12.67	-11.00	-4.50	27.4
curseforge	Amazon	19.1	+17.23	-11.00	-6.38	16.4
dpreview	Amazon	19.1	+8.54	-11.00	-4.14	35.1
futhead	Amazon	19.1	-99.31	-11.00	-15.89	5.5
gamepedia	Amazon	19.1	+33.74	-11.00	+1.94	28.9
goodreads	Amazon	19.1	-4.03	-11.00	-8.91	30.0
hearthpwn	Amazon	19.1	+6.13	-11.00	-10.54	2.7
imdb	Amazon	19.1	+28.74	-11.00	+20.47	79.2
twitch	Amazon	19.1	+46.31	-11.00	+17.86	50.4
bild	Axel Springer	25.6	+0.70	-8.78	-7.42	14.4
computerbild	Axel Springer	25.6	-134.17	-8.78	-13.15	3.5
lacentrale	Axel Springer	25.6	+3.32	-8.78	-0.67	67.1
logic-immo	Axel Springer	25.6	+12.83	-8.78	+0.48	42.8
seloger	Axel Springer	25.6	-6.06	-8.78	-6.80	72.9

Continued on next page

Table E.1 (continued)

Website family	Corporate parent	Ad share (%)	Raw synthetic	Model-implied	Empirical Bayes	Weight on raw (%)
welt	Axel Springer	25.6	-44.39	-8.78	-15.65	19.3
booking	Booking Holdings	7.3	+11.31	-15.02	+0.56	59.2
rentalcars	Booking Holdings	7.3	+7.03	-15.02	-10.95	18.4
hotels	Expedia	9.7	+18.03	-14.19	-0.69	41.9
bing	Microsoft	57.1	-2.28	+1.92	+1.15	18.4
linkedin	Microsoft	57.1	+16.67	+1.92	+7.92	40.7
live	Microsoft	57.1	-9.65	+1.92	-4.11	52.2
msn	Microsoft	57.1	-21.65	+1.92	-3.89	24.7
slideshare	Microsoft	57.1	-23.07	+1.92	-6.39	33.2
leboncoin	Schibsted	37.3	-25.81	-4.81	-10.12	25.3
milanuncios	Schibsted	37.3	-16.67	-4.81	-5.41	5.1
subito	Schibsted	37.3	-24.14	-4.81	-7.03	11.5
tripadvisor	TripAdvisor (borderline)	71.6	-0.02	+6.84	+4.31	36.9
wish	Wish (ContextLogic)	12.4	-68.36	-13.27	-35.83	41.0
2ememain	eBay	11.4	-35.10	-13.62	-17.37	17.4
ebay	eBay	11.4	+16.63	-13.62	+4.92	61.3
ebay-kleinanzeigen	eBay	11.4	-50.57	-13.62	-18.89	14.3
kijiji	eBay	11.4	-11.20	-13.62	-13.31	12.9
marktplaats	eBay	11.4	-80.02	-13.62	-25.95	18.6
mobile	eBay	11.4	-64.83	-13.62	-18.01	8.6
<i>Pure intermediation (5 families)</i>						
airbnb	Airbnb	0.0	+0.70	-17.50	-9.55	43.7
viki	Rakuten	0.0	+11.20	-17.50	-9.59	27.5
indeed	Recruit Holdings	0.0	+7.70	-17.50	-8.11	37.2
uber	Uber	0.0	+5.53	-17.50	-12.96	19.7
zalando	Zalando	0.0	-0.03	-17.50	-15.44	11.8

Notes: Advertising share is 100C. The three effect columns report log percentage points, without exponentiation; the final column reports the percentage weight on the raw contrast in the EB estimate. The displayed point estimates use the original fitted synthetic paths. The AR(1) noise calibration is a working within-website variance model; its formula does not add a separate donor-weight variance component. In the full-pipeline bootstrap, synthetic weights, model centers and this calibration are all reestimated, so the reported group inference incorporates fitted-weight and shared-donor variation. The table reports point estimates; reliability weights describe shrinkage and are not confidence levels.

F Supplementary heterogeneity analyses

The comparison groups follow the main-text exposure definition. Exposure requires a positive French-panel detection during September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The residual group includes unknown exposure. The 638 nonliable families divide into 469 exposed and 169 residual families.

Bootstrap inference for supplementary specifications. All standard errors and plotted intervals in this appendix use 1,999 draws with the same ordinary parent and exposure-stratified nonliable-family resampling as the main regression exhibits. Each draw retains full histories and original missingness, reestimates donor-change scales and baseline offsets, and fits synthetic weights on available June–November 2018 French observations. The French-share specifications repeat the original $\kappa = 1$ share-gap penalty and rebuild each synthetic member’s weighted donor-share regressor; the joint revenue-share specification instead refits the headline traffic-only weights without this penalty. Donor copies receive separate simplex coordinates and ridge penalties, with shared multiplicities across all targets and pools. Original 55-family standardization, French-share calibration, exposure labels and eligibility rules remain fixed. Both annual specifications and primary monthly regressions are reestimated. Standard errors use the draw standard deviation, and plotted intervals are pointwise bootstrap percentiles. Appendix H reports the conditional results.

F.1 French share of global traffic

In addition to the baseline regression, I examine whether traffic responses vary with French audience concentration. Equation 2 motivates this comparison: the effective marginal tax rate depends on French presence and French advertising prices relative to the worldwide average. I measure audience concentration using family-level French and global WTM popularity over September–November 2018. To account for separate geographic normalization, I multiply each month’s French popularity by 0.9 times Amazon.fr’s global-to-French popularity ratio, sum across available months and divide by the family’s global popularity summed over the same months. This calibration assumes Amazon.fr obtains 90% of its traffic from France. Numerator and denominator use identical months with positive French and global observations. The measure is held fixed thereafter, without clipping, and standardized using the original 55 liable families’ mean (16.28%) and SD (26.00 percentage points).

The equal-weighted specification is

$$y_{it} = \alpha_i + \lambda_t + \beta(L_i \times Post_t) + \delta(\tilde{F}_i \times Post_t) + \gamma(L_i \times \tilde{F}_i \times Post_t) + \varepsilon_{it}. \quad (26)$$

Every actual liable and control family uses its own standardized share, $\tilde{F}_i$. The control gradient is δ , the liable gradient is $\delta + \gamma$, and γ measures their difference per original SD. The coefficient β compares the

groups when both have the original liable-sample mean French share. The period, equal family-month weights and parent/control-family resampling blocks follow Equation 8.

For comparability with the business-model heterogeneity regressions, I also estimate the liable-share-only difference-in-differences specification

$$y_{it} = \alpha_i + \lambda_t + \beta_D(L_i \times Post_t) + \theta_D(L_i \times Post_t \times \tilde{F}_i) + \varepsilon_{it}. \quad (27)$$

This benchmark omits the share-by-post term for controls, constraining their exposure gradient to zero. I estimate it both with actual controls and on the same preferred synthetic pairs. It is distinct from the own-member-share specification in Equation 26.

For the preferred synthetic specification, I stack each liable family and its synthetic comparison and estimate the interactions in Equation 26 with member fixed effects and shared month fixed effects. The liable member uses its own share; the synthetic member uses the weighted average of its donors' shares, using the same weights as the traffic outcome. For this heterogeneity analysis, I augment the headline French-only traffic-fitting criterion with $(\tilde{F}_i - \sum_j w_{ij} \tilde{F}_j)^2$, a soft French-share-balance penalty with coefficient $\kappa = 1$. Demeaning, donor-change noise scaling and ridge coefficient $\sqrt{12}$ follow the headline fit. Donors require complete French June–November 2018 and January–December 2019 observations, with no German, US or global coverage restriction. Each target can use all 159 residual, 462 exposed or 621 pooled donors in the selected pool. This additional criterion improves French-share balance without imposing equality. Mean absolute share gaps are 2.45, 1.64 and 1.45 percentage points for residual, advertising-exposed and pooled donors, respectively. All 55 liable families remain in the analysis. The French-share penalty is specific to this heterogeneity analysis; headline average effects and joint revenue-share regressions use French-only traffic weights without it.

The monthly version replaces the three post interactions with month-specific interactions. I estimate June–November 2018 and January–December 2019 jointly and recenter the coefficients on September–November 2018 in every bootstrap draw, retaining the cross-month dependence in the resulting draws. A causal interpretation of γ requires that, absent the tax, the relationship between traffic changes and French share would have been the same in the liable and comparison groups, as well as valid synthetic counterfactuals. Prepolicy patterns and imperfect exposure balance remain relevant diagnostics.

Heterogeneity by French share of global traffic. Figure F.1 separates synthetic controls on the left from equal-weighted actual controls on the right, using unexposed controls in both columns. The upper panels show fitted liable and comparison paths at hypothetical French traffic shares of 0% and 100%, using the original sample mean and standard deviation to standardize the share. The lower panels display the own-member-share differential gradient γ_m and the liable-share-only DiD gradient $\theta_{D,m}$ for each comparison method. Table F.1, Panel A, reports the period coefficients β , δ and γ , together with the liable gradient $\delta + \gamma$. Panel B reports the liable-share-only DiD coefficients β_D and θ_D . The table retains all three control groups

and reports full-pipeline bootstrap standard errors, sample counts and fit statistics for both specifications. French shares use the Amazon.fr calibration conditional on its assumed 90% French traffic share.

[Author: add interpretation of Figure F.1 and Table F.1.]

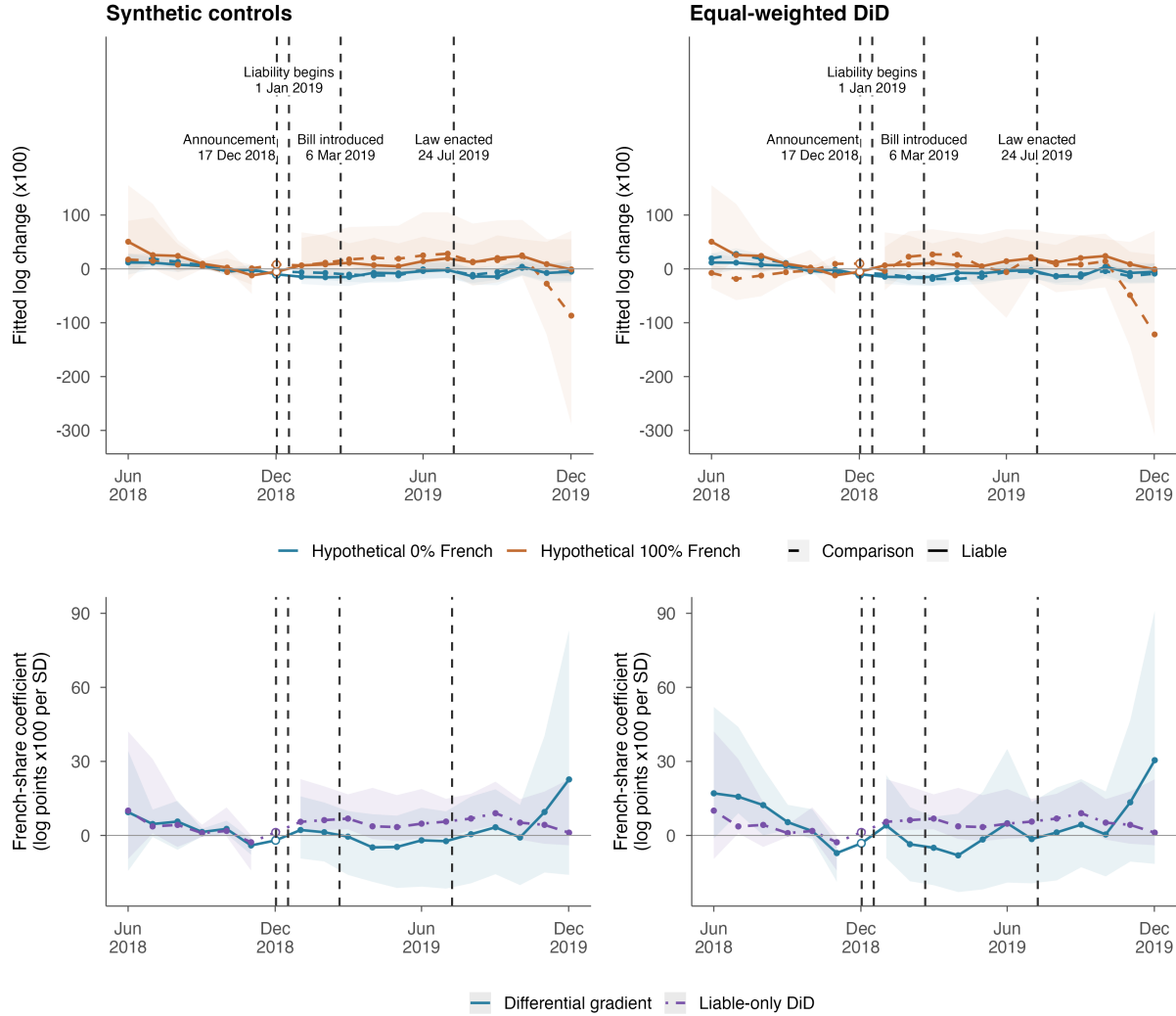


Figure F.1: French-share heterogeneity with traffic-and-exposure-matched synthetic controls

Notes: Left: stacked synthetic controls; right: equal-weighted actual controls, using the residual comparison group. Upper panels show fitted liable and comparison paths at hypothetical 0% and 100% French audience shares, using the original share standardization. Lower panels show the own-member-share differential gradient and the liable-share-only DiD gradient. All paths are centered on September–November 2018 and measured as 100 times log units; gradients are per original SD. Shading gives pointwise 95% percentile intervals from 1,999 full-history parent/family draws. Every draw refits French-only synthetic scaling and weights with the $\kappa = 1$ share-gap penalty, rebuilds outcome-weighted donor shares and baseline offsets, and reestimates both regressions. Donor multiplicities are shared across targets. December is display-only without an interval. Appendix H preserves the conditional figure.

Table F.1: French-share heterogeneity: own-share and DiD specifications

Model Estimator	No observed advertising exposure		Advertising-exposed controls		Pooled controls	
	(1) Actual	(2) Stacked SC	(3) Actual	(4) Stacked SC	(5) Actual	(6) Stacked SC
<i>Panel A: Own-member-share specification</i>						
Liabile $\times$ Post (β)	+5.37 (7.83)	−0.90 (7.88)	+8.43 (7.10)	+3.55 (7.47)	+7.56 (6.89)	+3.26 (7.03)
Share $\times$ Post (δ)	+1.96 (7.22)	+3.17 (7.41)	+1.56 (2.28)	+2.19 (4.52)	+1.32 (2.13)	+2.33 (4.34)
Liabile $\times$ Share $\times$ Post (γ)	+3.25 (9.56)	+2.03 (8.52)	+3.64 (6.98)	+3.01 (8.41)	+3.89 (6.89)	+2.88 (7.69)
Liabile share gradient ($\delta + \gamma$)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)
<i>Panel B: Liabile-share-only DiD</i>						
Liabile $\times$ Post (β_D)	+5.74 (7.82)	−0.91 (8.12)	+7.52 (7.06)	+3.46 (7.57)	+7.06 (6.88)	+3.19 (7.20)
Liabile $\times$ Share $\times$ Post (θ_D)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)	+5.20 (6.58)
Liabile families	55	55	55	55	55	55
Liabile parent firms	20	20	20	20	20	20
Raw comparison families	169	—	469	—	638	—
Synthetic comparison units	—	55	—	55	—	55
Eligible donor pool	—	159	—	462	—	621
Eligible donors per liable family	—	159	—	462	—	621
Regression units	224	110	524	110	693	110
Observations	3,311	1,650	7,845	1,650	10,331	1,650
Liabile / nonliable resampling blocks	20 / 169	20 / 169	20 / 469	20 / 469	20 / 638	20 / 638
Unit fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (A)	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2 (A)	0.00165	0.00818	0.00086	0.00918	0.00063	0.00903
RMSE (log units) (A)	0.369	0.182	0.480	0.178	0.462	0.179
R^2 (B)	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2 (B)	0.00150	0.00647	0.00056	0.00814	0.00043	0.00786
RMSE (log units) (B)	0.369	0.182	0.480	0.178	0.462	0.179

Notes: September–November 2018 versus January–December 2019; December excluded. Coefficients and standard errors are multiplied by 100. Panel A estimates Equation (26); Panel B estimates Equation (27). French shares retain the original 55-family mean (16.28%) and SD (26.00 percentage points), using the fixed Amazon.fr 90% calibration. Synthetic members use outcome-weighted donor shares. In every draw, French-only prepolicy fitting retains the squared standardized share-gap penalty $\kappa = 1$, nonnegative unit-sum weights and ridge coefficient $\sqrt{12}$. All 159, 462 or 621 original eligible donors can enter their resampled pools; no foreign/global coverage restriction is added. The synthetic share regressor is reconstructed from the newly fitted weights. All standard errors are sample standard deviations from 1,999 full-history pairs-bootstrap draws. Each draw samples 20 whole liable parents and samples nonliable families separately within the 469 exposed and 169 residual groups. Family histories and missingness remain together; multiplicities are shared across target and donor roles. All donor-change scales, synthetic weights, baseline offsets and regression coefficients are reestimated. Repeated donor copies receive separate simplex coordinates and ridge penalties. Exposure definitions, original standardization, sample counts and original-sample fit statistics are unchanged. Pointwise intervals use bootstrap percentiles. Appendix H reports the conditional exhibits.

F.2 Joint advertising and intermediation revenue shares

The joint specification includes advertising and intermediation revenue separately, each expressed as a share of total consolidated revenue:

$$y_{it} = \alpha_i + \lambda_t + \beta (L_i \times Post_t) + \theta_A (L_i \times Post_t \times \tilde{A}_{j(i)}) + \theta_I (L_i \times Post_t \times \tilde{I}_{j(i)}) + \varepsilon_{it}, \quad (28)$$

where $A_j = R_j^A/R_j$ and $I_j = R_j^I/R_j$, with R_j denoting total consolidated revenue. These shares need not sum to one because firms can earn revenue from other activities. They describe reliance on each revenue source rather than the exact fraction of revenue subject to the French tax.

The supplementary joint revenue-share specification uses the same baseline period, equal family-month weights and fixed effects as the main DiD. Revenue measures are held at fiscal-year 2018 values. Tildes subtract the original liable-family mean and divide by the original liable-family standard deviation. The coefficient β compares groups at mean advertising and intermediation shares; θ_A and θ_I are the respective gradients holding the other share fixed. Share interactions enter only for liable members, including in the stacked synthetic specification.

For Equation 28, I instead consider a firm earning all consolidated revenue from advertising, $(A_j, I_j) = (1, 0)$, and a firm earning all consolidated revenue from intermediation, $(A_j, I_j) = (0, 1)$. Their implied responses are

$$\Delta_A^{AI} = \beta + \theta_A \frac{1 - \bar{A}}{s_A} + \theta_I \frac{-\bar{I}}{s_I}, \quad (29)$$

$$\Delta_I^{AI} = \beta + \theta_A \frac{-\bar{A}}{s_A} + \theta_I \frac{1 - \bar{I}}{s_I}. \quad (30)$$

These profiles summarize the fitted relationships at economically interpretable revenue compositions. They are predictions for hypothetical firms and may involve extrapolation beyond the observed combinations of revenue shares. I evaluate each profile using the entire coefficient vector in every full-pipeline bootstrap draw. Standard errors are the standard deviation across draws, and confidence intervals are the 2.5th and 97.5th percentiles; this retains dependence among the estimated coefficients.

Advertising and intermediation shares of total revenue. Figure F.2 presents the monthly version of Equation (28), which enters advertising and intermediation shares of total revenue jointly. It compares stacked synthetic controls on the left with equal-weighted actual controls on the right, using controls without observed advertising exposure in both columns. Its upper panels display fitted liable and comparison paths for hypothetical advertising, intermediation and neither-activity firms, defined by $(A, I) = (1, 0)$, $(0, 1)$ and $(0, 0)$. Its lower panels display the corresponding liable-minus-comparison contrasts. Each contrast combines β_m , $\theta_{A,m}$ and $\theta_{I,m}$ evaluated at the stated profile using the original share standardization.

[Author: add interpretation of Figure F.2 and the joint revenue-share specification in Table F.2.]

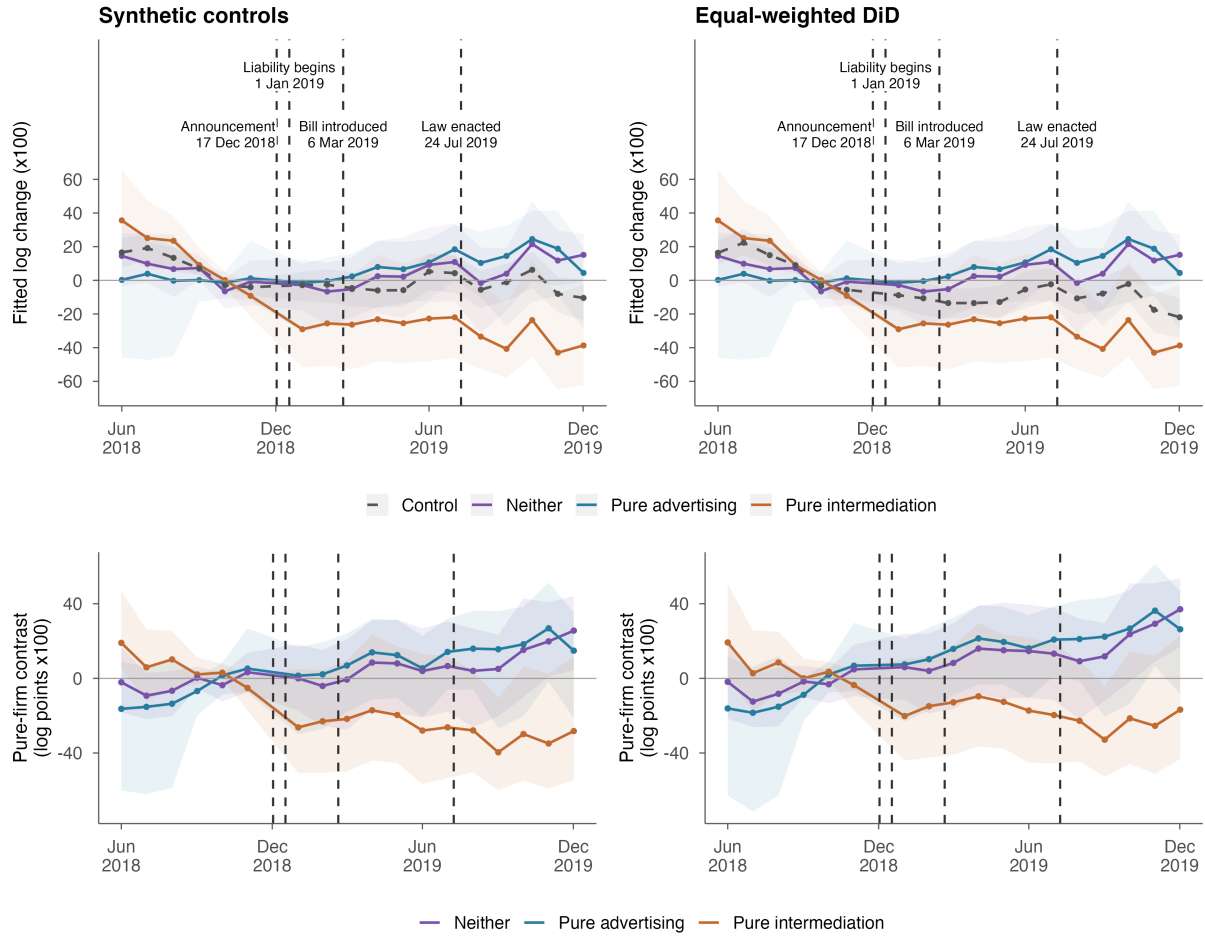


Figure F.2: Advertising and intermediation revenue shares and traffic

Notes: Left: stacked synthetic controls; right: equal-weighted actual controls, using the residual comparison group. Upper panels show fitted paths for hypothetical advertising, intermediation and neither-activity firms, $(A, I) = (1, 0), (0, 1), (0, 0)$, and the comparison path; lower panels show their contrasts. Shares retain original liable-sample standardization. Paths are centered on September–November 2018 in 100 times log units; December is excluded. Shading gives pointwise 95% percentile intervals from 1,999 whole-parent and nonliable-family draws. Every draw refits French-only synthetic scales, traffic-only weights, baseline offsets and the joint revenue-share regression. Original weights do not have a French-share penalty. Shared donor draws retain dependence across targets. Appendix H preserves the conditional figure.

Table F.2 reports β , θ_A and θ_I from Equation (28) for actual-control and stacked synthetic DiD, retaining unexposed, advertising-exposed and pooled controls. Revenue interactions enter only for liable members. Full-pipeline bootstrap standard errors, fit statistics, sample counts and fixed-effect indicators accompany the coefficients. The period comparison is September–November 2018 versus January–December 2019, excluding December 2018. Coefficients are multiplied by 100; share gradients are per original liable-sample SD.

[Author: add discussion of the joint revenue-share estimates in Table F.2.]

Table F.2: Joint advertising and intermediation revenue-share heterogeneity

Estimator Model	No observed advertising exposure		Advertising-exposed controls		Pooled controls	
	Actual DiD (1)	Stacked SC (2)	Actual DiD (3)	Stacked SC (4)	Actual DiD (5)	Stacked SC (6)
Liabe $\times$ Post (β)	+5.74 (7.21)	−2.81 (7.35)	+7.52 (6.25)	+3.75 (6.55)	+7.06 (6.08)	+2.58 (6.24)
Liabe $\times$ Post $\times$ Advertising (θ_A)	+1.23 (5.47)	+1.23 (5.47)	+1.23 (5.47)	+1.23 (5.47)	+1.23 (5.47)	+1.23 (5.47)
Liabe $\times$ Post $\times$ Intermediation (θ_I)	−11.69 (6.37)	−11.69 (6.37)	−11.69 (6.37)	−11.69 (6.37)	−11.69 (6.37)	−11.69 (6.37)
R^2	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2	0.00489	0.03743	0.00141	0.03734	0.00112	0.03662
RMSE (log units)	0.368	0.174	0.480	0.176	0.462	0.175
Liabe families / parents	55 / 20	55 / 20	55 / 20	55 / 20	55 / 20	55 / 20
Raw comparison families	169	—	469	—	638	—
Synthetic comparison units	—	55	—	55	—	55
Eligible donor pool	—	159	—	462	—	621
Eligible donors per liabe family	—	159	—	462	—	621
Observations	3,311	1,650	7,845	1,650	10,331	1,650
Regression units	224	110	524	110	693	110
Liabe / nonliabe resampling blocks	20 / 169	20 / 169	20 / 469	20 / 469	20 / 638	20 / 638
Unit fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Equation (28); September–November 2018 versus January–December 2019, excluding December. Coefficients and standard errors are multiplied by 100; gradients are per original liabe-sample SD of the fiscal-year 2018 revenue shares. Actual regressions use observed families; stacked comparisons use one liabe and one synthetic member per target-month. Synthetic fitting uses French prepolicy traffic, donor-change scaling, nonnegative unit-sum weights and ridge $\sqrt{12}$, without a French-share penalty. The eligible pools contain 159, 462 and 621 French-complete families. All standard errors are sample standard deviations from 1,999 full-history pairs-bootstrap draws. Each draw samples 20 whole liabe parents and samples nonliabe families separately within the 469 exposed and 169 residual groups. Family histories and missingness remain together; multiplicities are shared across target and donor roles. All donor-change scales, synthetic weights, baseline offsets and regression coefficients are reestimated. Repeated donor copies receive separate simplex coordinates and ridge penalties. Exposure definitions, original standardization, sample counts and original-sample fit statistics are unchanged. Pointwise intervals use bootstrap percentiles. Appendix H reports the conditional exhibits.

G Alternative comparison groups

The comparison groups follow the main-text exposure definition. Exposure requires a positive French-panel detection during September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The residual group includes unknown exposure. The 638 nonliable families divide into 469 exposed and 169 residual families.

Synthetic comparisons use the current French-only traffic fit without French-share balancing. Every target can use all 462 exposed or 621 pooled donors; donor eligibility requires complete French June–November 2018 and January–December 2019 coverage without foreign/global restrictions.

Standard errors follow the main full-pipeline procedure: 1,999 draws resample 20 whole liable parents and nonliable families within exposure groups, preserving histories and missingness. Every draw refits the applicable French-only synthetic weights, donor-change scales, baseline offsets and regressions. Shared donor multiplicities preserve dependence across targets; original composition standardization is retained. Appendix H reports the conditional tables.

G.1 Headline traffic estimates

Table G.1 reports the headline DiD specifications using advertising-exposed and pooled controls. Each control group has an equal-weighted actual-control column and a stacked synthetic-control column. Coefficients, full-pipeline bootstrap standard errors, sample counts and fit statistics use the same specifications and estimation periods as Table 7, which reports the unexposed comparison in the main text.

[Author: add interpretation of the alternative-control comparisons in Table G.1.]

Table G.1: Headline traffic estimates using advertising-exposed and pooled controls

Model Estimator Dependent variable	Advertising-exposed controls		Pooled controls	
	(1)	(2)	(3)	(4)
	DiD y_{it}	Stacked SC y_{ist}	DiD y_{it}	Stacked SC y_{ist}
Liabie $\times$ Post (β)	+7.52 (6.82)	+3.75 (7.11)	+7.06 (6.62)	+2.58 (6.76)
Liabie families	55	55	55	55
Liabie parent firms	20	20	20	20
Raw comparison families	469	–	638	–
Synthetic comparison units	–	55	–	55
Eligible donor pool	–	462	–	621
Eligible donors per liabie family	–	462	–	621
Regression units	524	110	693	110
Observations	7,845	1,650	10,331	1,650
Liabie / nonliabie resampling blocks	20 / 469	20 / 469	20 / 638	20 / 638
Unit fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
R^2	0.911	0.994	0.914	0.994
Within R^2	0.00037	0.00175	0.00027	0.00084
RMSE (log units)	0.480	0.179	0.462	0.178

Notes: January–December 2019 versus September–November 2018; December excluded. Coefficients are 100β . Family/member and month fixed effects, equal family-month weights and the French-only donor eligibility and fitting criterion follow the main specification. Exposed and pooled synthetic donor pools contain 462 and 621 eligible families. All standard errors are sample standard deviations from 1,999 full-history pairs-bootstrap draws. Each draw samples 20 whole liabie parents and samples nonliabie families separately within the 469 exposed and 169 residual groups. Family histories and missingness remain together; multiplicities are shared across target and donor roles. All donor-change scales, synthetic weights, baseline offsets and regression coefficients are reestimated. Repeated donor copies receive separate simplex coordinates and ridge penalties. Exposure definitions, original standardization, sample counts and original-sample fit statistics are unchanged. Pointwise intervals use bootstrap percentiles. Appendix H reports the conditional exhibits.

G.2 Advertising composition of in-scope revenue

Table G.2 reports β and θ_C from Equation (12), using advertising-exposed and pooled control groups. Actual-control and stacked synthetic DiD columns include full-pipeline bootstrap standard errors, sample counts and fit statistics. Table 9 reports the unexposed comparison in the main text.

[Author: add interpretation of the alternative-control comparisons in Table G.2.]

Table G.2: Advertising-composition estimates using advertising-exposed and pooled controls

Estimator Model	Advertising-exposed controls		Pooled controls	
	Actual DiD (1)	Stacked SC (2)	Actual DiD (3)	Stacked SC (4)
Liabe $\times$ Post (β)	+7.52 (5.82)	+3.75 (6.01)	+7.06 (5.60)	+2.58 (5.65)
Liabe $\times$ Post $\times$ Composition (θ_C)	+12.38 (4.92)	+12.38 (4.92)	+12.38 (4.92)	+12.38 (4.92)
R^2	0.911	0.994	0.914	0.994
Within R^2	0.00147	0.03934	0.00117	0.03864
RMSE (log units)	0.480	0.175	0.462	0.175
Liabe families / parents	55 / 20	55 / 20	55 / 20	55 / 20
Raw comparison families	469	–	638	–
Synthetic comparison units	–	55	–	55
Observations	7,845	1,650	10,331	1,650
Regression units	524	110	693	110
Liabe / nonliabe resampling blocks	20 / 469	20 / 469	20 / 638	20 / 638
Unit fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes

Notes: Equation (12); September–November 2018 versus January–December 2019, excluding December. Coefficients and standard errors are multiplied by 100. Composition retains its original 55-family mean and SD. Actual-control and stacked specifications use equal family-month weights and the headline French-only fitting criterion without French-share balancing. Eligible exposed and pooled donor sets contain 462 and 621 families. All standard errors are sample standard deviations from 1,999 full-history pairs-bootstrap draws. Each draw samples 20 whole liabe parents and samples nonliabe families separately within the 469 exposed and 169 residual groups. Family histories and missingness remain together; multiplicities are shared across target and donor roles. All donor-change scales, synthetic weights, baseline offsets and regression coefficients are reestimated. Repeated donor copies receive separate simplex coordinates and ridge penalties. Exposure definitions, original standardization, sample counts and original-sample fit statistics are unchanged. Pointwise intervals use bootstrap percentiles. Appendix H reports the conditional exhibits.

H Sensitivity to conditioning on fitted synthetic controls

This appendix reports inference that conditions on fitted synthetic controls. Point estimates, outcomes, donor eligibility, exposure definitions and samples are identical to the main analysis. Only the uncertainty calculation differs. The regression tables and figures use parent-clustered inference for liable families and family-clustered inference for nonliable families, treating the fitted synthetic paths as given. Raw pure-group and empirical Bayes (EB) averages use revenue-stratified whole-parent draws with synthetic weights held fixed: 4,999 draws for annual effects and 1,999 for monthly paths. Each fixed-weight EB draw reestimates model centers, temporal noise and shrinkage parameters.

The main text instead uses 1,999 draws that resample nonliable families as well as liable parents and reconstruct synthetic comparisons in every draw. Ordinary parent draws are used for regressions and model-only profiles; the raw pure-group and EB averages retain the revenue-stratified parent design. Table H.1 reports both the conditional standard errors and fixed-weight bootstrap standard errors calculated on the same draws as the refitted-weight procedure. The paired comparison isolates the effect of donor resampling and weight refitting from the change in inferential method. The variance change need not equal a separate additive first-stage variance, because the paired estimates are dependent.

The composition gradient is identical under fixed and refitted weights within a draw: in the stacked model the synthetic members enter the common comparison intercept, while the composition slope is determined by the liable members.

Table H.1: Annual inference with fixed and refitted synthetic weights

Quantity	Conditional SE	Paired fixed SE	Refitted SE
Liable vs residual	6.38	6.18	7.78
Exposed vs residual	2.98	2.94	4.96
Composition intercept	5.28	5.15	6.99
Composition gradient	4.96	4.92	4.92
Pure advertising / Composition model	7.57	7.75	9.11
Pure advertising / Raw group average	6.60	6.72	7.97
Pure advertising / Empirical Bayes	6.41	6.66	7.81
Pure intermediation / Composition model	9.26	8.94	10.09
Pure intermediation / Raw group average	1.87	1.88	6.59
Pure intermediation / Empirical Bayes	6.35	6.51	8.24

Notes: All standard errors are in log percentage points. Conditional SEs use the clustered or fixed-weight bootstrap methods described above; paired fixed and refitted SEs use identical target draws. Regressions and model endpoints use ordinary draws of 20 whole liable parents, while the raw and EB pure-group averages use revenue-stratified draws of 5 advertising, 10 mixed and 5 intermediation parents. Nonliable families are sampled within the 469 exposed and 169 residual families, sharing multiplicities across target and donor roles. The refitted arm repeats donor eligibility, scaling, fitting, baseline alignment, regressions and EB calculations. The fixed arm retains saved synthetic paths. Both arms use 1,999 draws. Analytical clustered SEs and the 4,999-draw annual EB calculation are not numerically identical to the paired fixed-weight bootstrap. Only five parents belong to each pure group.

Table H.2: Fixed-weight sensitivity: difference-in-differences estimates of French traffic responses

Model Estimator Dependent variable	No documented advertising exposure	
	(1) DiD y_{it}	(2) Stacked SC y_{ist}
Liabe $\times$ Post (β)	+5.74 (7.71)	−2.81 (6.38)
Liabe families	55	55
Liabe parent firms	20	20
Raw comparison families	169	–
Synthetic comparison units	–	55
Eligible donor pool	–	159
Eligible donors per liabe family	–	159
Regression units	224	110
Observations	3,311	1,650
Clusters	189	20
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
R^2	0.953	0.994
Within R^2	0.00073	0.00101
RMSE (log units)	0.369	0.177

Notes: January–December 2019 versus September–November 2018; December 2018 excluded. Coefficients are 100β , with clustered standard errors in parentheses. Column (1) estimates Equation (8) on observed families. Column (2) estimates $y_{ist} = \alpha_{is} + \lambda_t + \beta(L_s \times Post_t) + \varepsilon_{ist}$ on one actual liabe member and one synthetic member per family-month. Raw columns cluster by liabe parent or nonliabe family; synthetic pairs share the liabe parent cluster. Synthetic inference conditions on fitted weights and omits fitted-weight/shared-donor uncertainty.

Table H.3: Fixed-weight sensitivity: difference-in-differences estimates for advertising-exposed nonliable sites

Model	(1)	(2)
Estimator	Conventional DiD	Stacked synthetic DiD
Dependent variable	y_{it}	y_{ist}
Exposed $\times$ Post (δ)	−1.75 (5.52)	−8.24 (2.98)
Exposed nonliable families	469	469
Raw unexposed families	169	–
Synthetic comparison units	–	469
Starting eligible donor pool	–	159
Eligible donors per exposed family	–	159
Regression units	638	938
Observations	9,506	14,040
Clusters	638	469
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
R^2	0.895	0.934
Within R^2	0.00004	0.00212
RMSE (log units)	0.477	0.358

Notes: January–December 2019 versus September–November 2018; December 2018 excluded. Coefficients are 100δ (log points times 100); clustered standard errors in parentheses. Column (1) estimates $y_{it} = \alpha_i + \lambda_t + \delta(E_i \times Post_t) + \varepsilon_{it}$ on actual nonliable families, as in Equation (9). Column (2) stacks each exposed family with its own synthetic comparison and estimates $y_{ist} = \alpha_{is} + \lambda_t + \delta(T_s \times Post_t) + u_{ist}$, where $T_s = 1$ for the actual exposed member. Raw standard errors cluster by website family; both members of each synthetic pair share the exposed-family cluster. Synthetic inference conditions on fitted weights and omits fitted-weight/shared-donor uncertainty.

Table H.4: Fixed-weight sensitivity: traffic responses by advertising composition of in-scope revenue

Estimator Model	No documented advertising exposure	
	Actual DiD (1)	Stacked SC (2)
Liabe $\times$ Post (β)	+5.74 (6.82)	−2.81 (5.28)
Liabe $\times$ Post $\times$ Composition (θ_C)	+12.38 (4.84)	+12.38 (4.96)
R^2	0.953	0.994
Within R^2	0.00513	0.03948
RMSE (log units)	0.368	0.173
Liabe families / parents	55 / 20	55 / 20
Raw comparison families	169	–
Synthetic comparison units	–	55
Observations	3,311	1,650
Regression units	224	110
Clusters	189	20
Unit fixed effects	Yes	Yes
Month fixed effects	Yes	Yes

Notes: Estimates Equation (12). September–November 2018 versus January–December 2019, excluding December 2018. All coefficients and standard errors are multiplied by 100. Advertising composition uses fiscal-year 2018 accounts and is standardized at the mean and SD across the same 55 liable families. β is the liable-minus-comparison change at mean advertising composition; gradients are per liable-sample SD. Actual regressions cluster by liable parent or control family; synthetic pairs share their liable-parent cluster (20 clusters). Synthetic intervals condition on weights and omit fitted-weight/shared-donor uncertainty.

Table H.5: Fixed-weight sensitivity: pre/post traffic effects by business model and estimation method

	Composition model	Raw group average	Empirical Bayes
Pure advertising	+16.48 (7.57)	+22.25 (6.60)	+16.90 (6.41)
Pure intermediation	−17.50 (9.26)	+5.02 (1.87)	−11.13 (6.35)

Notes: Entries are annual pre/post traffic effects in log percentage points, with standard errors in parentheses. All effects compare September–November 2018 with January–December 2019; December 2018 is excluded. The composition-model column evaluates the original stacked model at advertising shares $C = 1$ and $C = 0$, using 55 liable website families under 20 parents. The raw column averages website-specific changes in observed-minus-synthetic log traffic among 13 observed pure advertising families and 5 observed pure intermediation families. The EB column averages their empirical Bayes estimates centered on predictions from the original stacked model. Raw and EB averages give equal weight to websites within each group. Model standard errors use the parent-clustered covariance matrix. Raw and EB standard errors use 4,999 revenue-group-stratified whole-parent bootstrap draws; each EB draw reestimates model predictions, noise parameters and residual signal variance. Inference conditions on fitted synthetic weights and omits fitted-weight/shared-donor uncertainty.

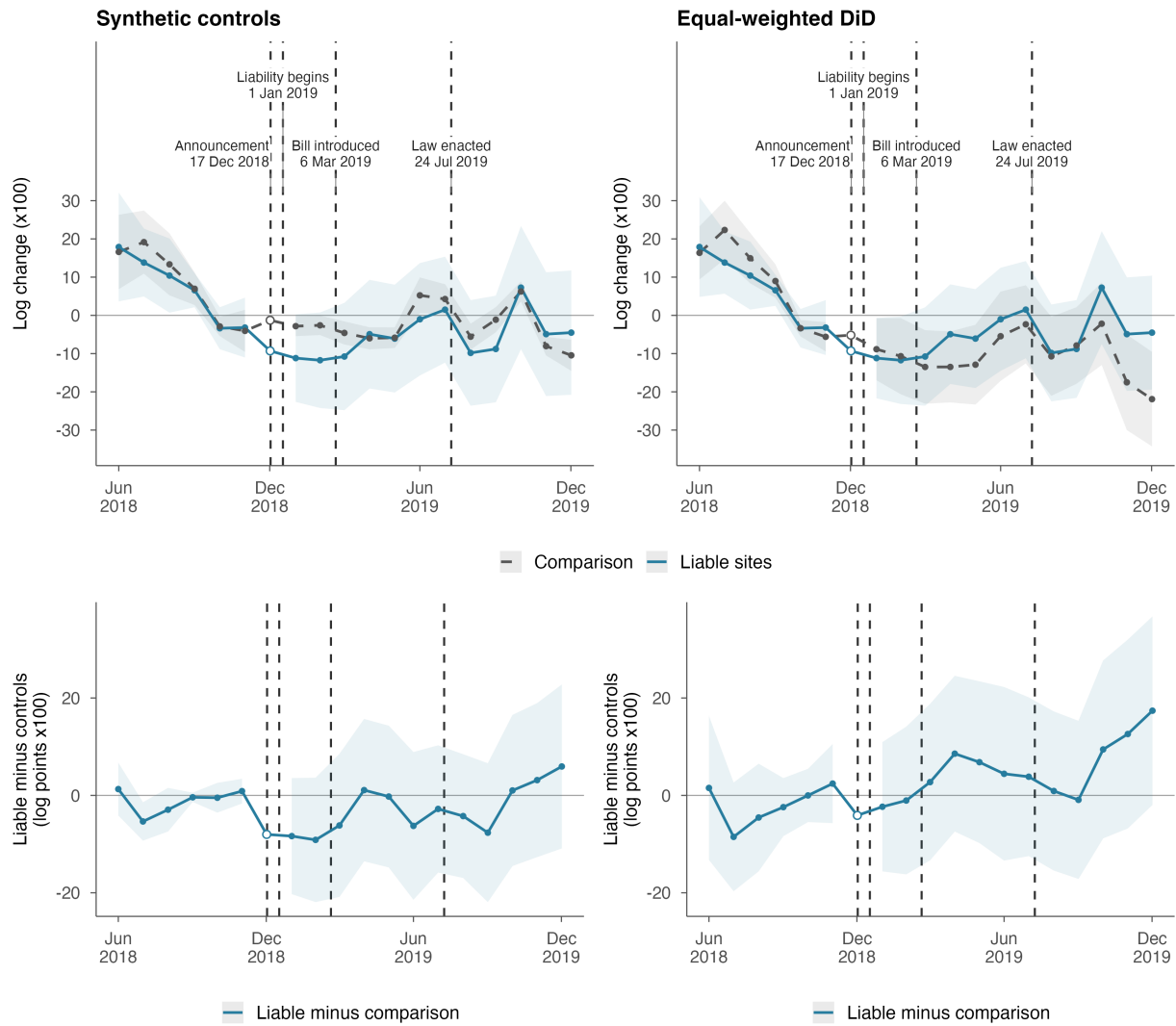


Figure H.1: Fixed-weight sensitivity: french traffic responses of liable platforms

Notes: Left: synthetic controls; right: equal-weighted actual-control DiD. Both compare the same 55 liable families under 20 parents with controls without documented advertising exposure. The raw pool contains 169 families; the starting eligible synthetic donor pool contains 159, with all 159 eligible donors available to every liable family. Upper panels show liable (blue solid) and comparison (gray dashed) paths; lower panels show liable-minus-comparison contrasts. Synthetic weights fit available June–November 2018 French traffic only. All series are centered on September–November 2018 and displayed as 100 times log changes or differences. Scales are common within rows; monthly points lie at calendar midpoints. Shading gives pointwise 95% intervals, clustered by liable parent for synthetic pairs and by liable parent or nonliable family for actual controls; synthetic intervals condition on fitted weights and omit weight/shared-donor uncertainty.

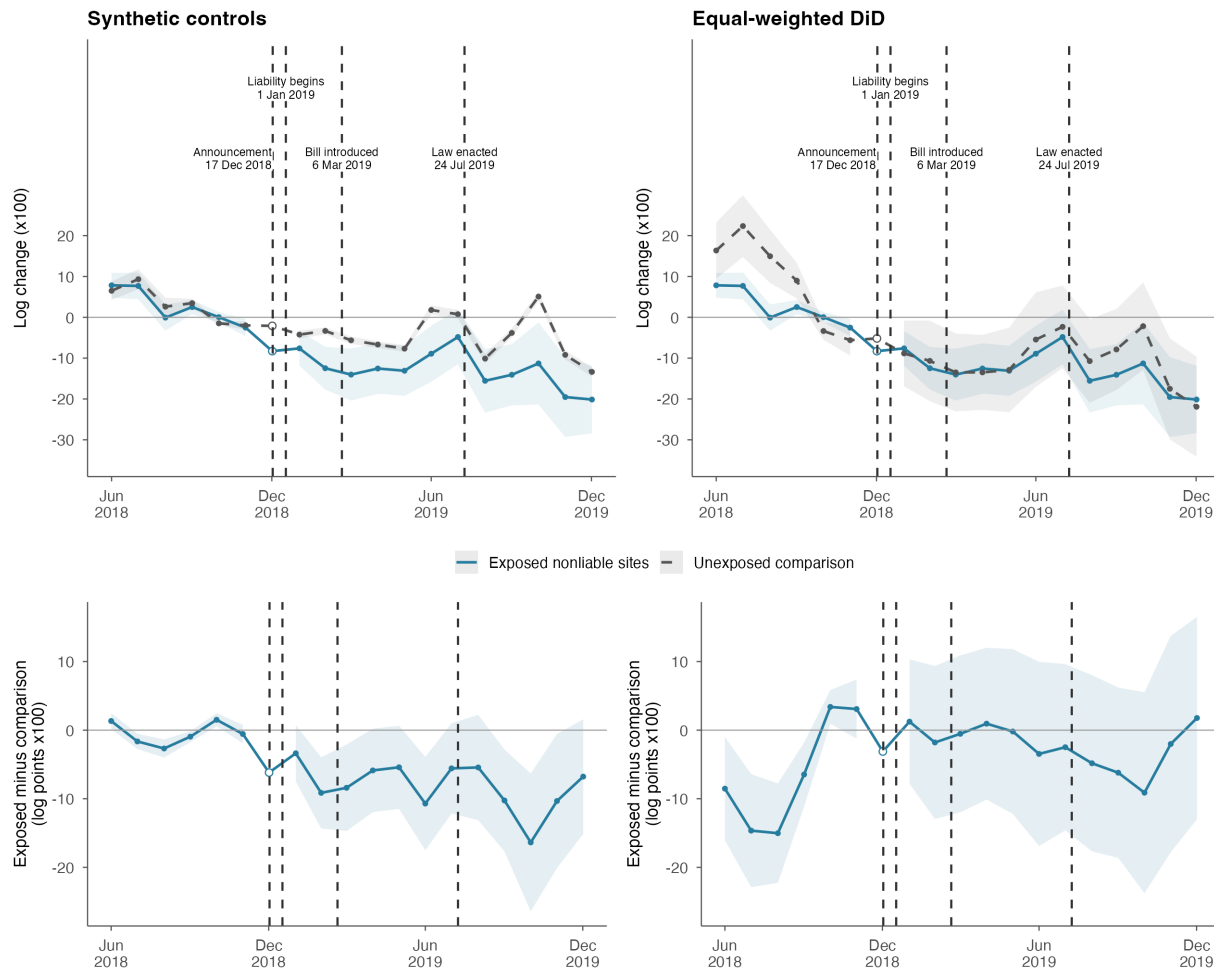


Figure H.2: Fixed-weight sensitivity: french traffic responses of advertising-exposed nonliable sites

Notes: The 469 exposed nonliable website families are compared with 169 families without documented advertising exposure and with individual synthetic comparisons drawn from a starting pool of 159 eligible donors, with all 159 donors eligible for every exposed family. Left column: individual synthetic controls. Right column: equal-weighted actual unexposed controls. Upper row: exposed and comparison paths. Lower row: the corresponding exposed-minus-comparison differences. Monthly outcomes are placed at calendar-month midpoints; connecting lines do not imply daily measurement. Paths are centered on September–November 2018 and displayed as $100b$ for the normalized log change or log difference b . Shading shows pointwise 95% family-clustered intervals, with both members of each synthetic pair assigned to the exposed-family cluster. Synthetic inference conditions on fitted weights and omits fitted-weight/shared-donor uncertainty.

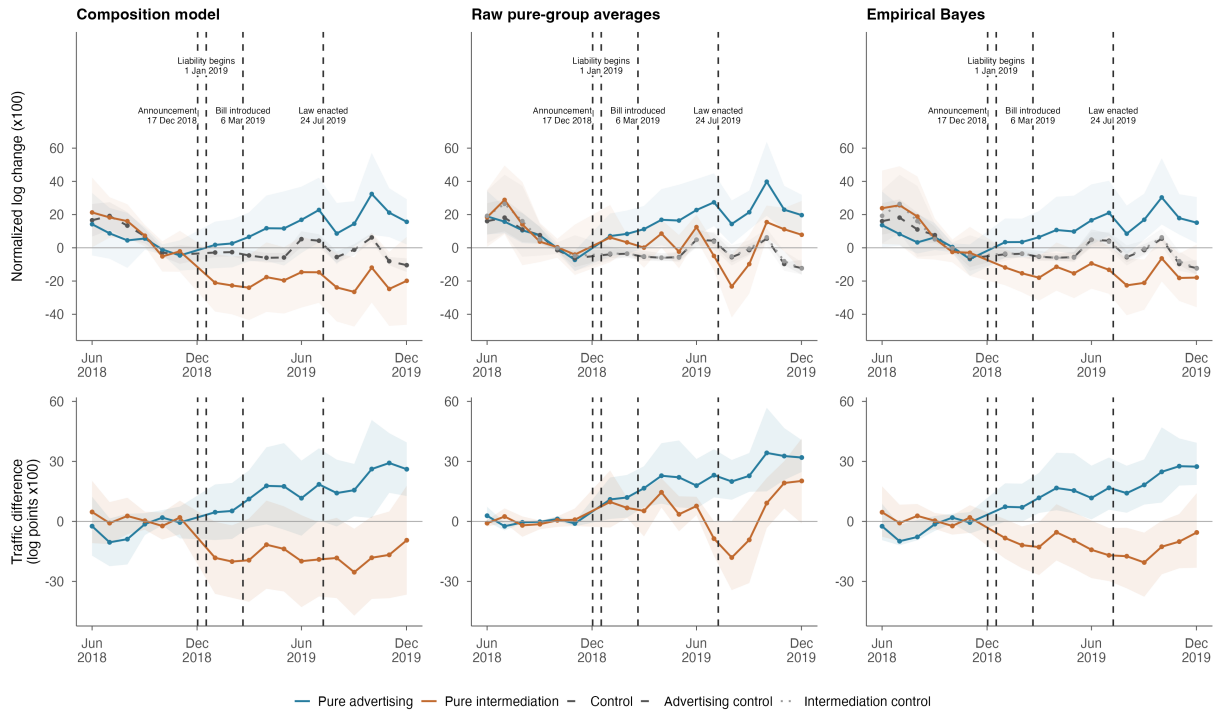


Figure H.3: Fixed-weight sensitivity: advertising composition and traffic

Notes: Left: hypothetical pure advertising ($C = 1$, blue) and pure intermediation ($C = 0$, orange) predictions from the original stacked composition model, estimated on 55 liable families under 20 parents. Middle: raw averages among 13 observed pure advertising families and 5 observed pure intermediation families. Right: empirical Bayes averages for those same groups, centered on predictions from the original model. Available families receive equal weight within each group. Upper panels show normalized traffic paths and synthetic comparisons; the model uses a common comparison path, while the raw and EB columns use each group's own average synthetic path. Lower panels show the corresponding traffic gaps. All paths are centered on September–November 2018 and measured as 100 times log changes or differences. Shading gives pointwise 95% intervals: parent-clustered t intervals for the model and percentile intervals from 1,999 revenue-group-stratified whole-parent draws for the raw and EB columns. Each EB draw reestimates original-model centers, noise parameters and residual signal variances. Inference conditions on fitted synthetic weights and omits fitted-weight/shared-donor uncertainty.

H.1 Conditional inference for the supplementary specifications

The following exhibits report fixed-weight supplementary analyses: the own-member French-share and liable-share-only specifications, joint revenue shares, and alternative comparison groups. Their coefficients and fitted paths are identical to the corresponding exhibits in Appendices [F](#) and [G](#); their standard errors and pointwise intervals condition on saved weights. The refitted-weight French-share analysis additionally reestimates the $\kappa = 1$ share-penalized weights and the outcome-weighted donor-share regressor in each bootstrap draw. Joint revenue-share and alternative-control analyses refit the headline traffic-only weights. The conditional exhibits below use parent/family-clustered inference as sensitivity checks.

Table H.6: Fixed-weight sensitivity: French-share heterogeneity: own-share and DiD specifications

Model Estimator	No observed advertising exposure		Advertising-exposed controls		Pooled controls	
	(1) Actual	(2) Stacked SC	(3) Actual	(4) Stacked SC	(5) Actual	(6) Stacked SC
<i>Panel A: Own-member-share specification</i>						
Liabe $\times$ Post (β)	+5.37 (7.79)	−0.90 (6.14)	+8.43 (7.00)	+3.55 (6.59)	+7.56 (6.78)	+3.26 (6.35)
Share $\times$ Post (δ)	+1.96 (6.64)	+3.17 (0.92)	+1.56 (2.21)	+2.19 (1.95)	+1.32 (2.07)	+2.33 (2.00)
Liabe $\times$ Share $\times$ Post (γ)	+3.25 (7.00)	+2.03 (2.26)	+3.64 (3.13)	+3.01 (3.19)	+3.89 (3.03)	+2.88 (3.13)
Liabe share gradient ($\delta + \gamma$)	+5.20 (2.22)	+5.20 (2.28)	+5.20 (2.21)	+5.20 (2.28)	+5.20 (2.21)	+5.20 (2.28)
<i>Panel B: Liabe-share-only DiD</i>						
Liabe $\times$ Post (β_D)	+5.74 (7.82)	−0.91 (6.31)	+7.52 (6.95)	+3.46 (6.70)	+7.06 (6.76)	+3.19 (6.47)
Liabe $\times$ Share $\times$ Post (θ_D)	+5.20 (2.22)	+5.20 (2.28)	+5.20 (2.21)	+5.20 (2.28)	+5.20 (2.21)	+5.20 (2.28)
Liabe families	55	55	55	55	55	55
Liabe parent firms	20	20	20	20	20	20
Raw comparison families	169	—	469	—	638	—
Synthetic comparison units	—	55	—	55	—	55
Eligible donor pool	—	159	—	462	—	621
Eligible donors per liabe family	—	159	—	462	—	621
Regression units	224	110	524	110	693	110
Observations	3,311	1,650	7,845	1,650	10,331	1,650
Clusters	189	20	489	20	658	20
Unit fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2 (A)	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2 (A)	0.00165	0.00818	0.00086	0.00918	0.00063	0.00903
RMSE (log units) (A)	0.369	0.182	0.480	0.178	0.462	0.179
R^2 (B)	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2 (B)	0.00150	0.00647	0.00056	0.00814	0.00043	0.00786
RMSE (log units) (B)	0.369	0.182	0.480	0.178	0.462	0.179

Notes: September–November 2018 versus January–December 2019; December 2018 excluded. Coefficients and clustered standard errors are multiplied by 100; share gradients are per original liabe-sample SD (26.00 percentage points). Panel A estimates Equation (26), and Panel B estimates Equation (27). Shares use matched pre-announcement French/global observations and the Amazon.fr calibration conditional on 90% French traffic, standardized at the original liabe mean (16.28%) and SD. Actual families use their own shares; synthetic members use outcome-weighted donor shares. Synthetic weights fit available June–November 2018 French traffic, with demeaning, donor-change noise scaling, ridge coefficient $\sqrt{12}$ and a squared standardized French-share-gap penalty ($\kappa = 1$). Donors require complete French June–November 2018 and January–December 2019 observations; there is no foreign/global coverage restriction. All targets have access to every donor in the selected pool; share balance remains approximate. Actual comparisons cluster by liabe parent/control family; synthetic pairs share the liabe-parent cluster. Synthetic inference conditions on fitted weights and omits fitted-weight/shared-donor uncertainty. Exposure requires a positive French-panel September–November 2018 detection on any family member domain of a study-designated liabe provider in the WTM advertising category, excluding adult advertising and custom overrides. The residual group includes unknown exposure. Fit statistics do not establish counterfactual validity.

Table H.7: Fixed-weight sensitivity: Joint advertising and intermediation revenue-share heterogeneity

Estimator Model	No observed advertising exposure		Advertising-exposed controls		Pooled controls	
	Actual DiD (1)	Stacked SC (2)	Actual DiD (3)	Stacked SC (4)	Actual DiD (5)	Stacked SC (6)
Liable $\times$ Post (β)	+5.74 (6.59)	−2.81 (4.72)	+7.52 (5.53)	+3.75 (4.81)	+7.06 (5.29)	+2.58 (4.68)
Liable $\times$ Post $\times$ Advertising (θ_A)	+1.23 (2.92)	+1.23 (3.00)	+1.23 (2.91)	+1.23 (3.00)	+1.23 (2.91)	+1.23 (3.00)
Liable $\times$ Post $\times$ Intermediation (θ_I)	−11.69 (5.75)	−11.69 (5.90)	−11.69 (5.73)	−11.69 (5.90)	−11.69 (5.73)	−11.69 (5.90)
R^2	0.953	0.994	0.911	0.994	0.914	0.994
Within R^2	0.00489	0.03743	0.00141	0.03734	0.00112	0.03662
RMSE (log units)	0.368	0.174	0.480	0.176	0.462	0.175
Liable families / parents	55 / 20	55 / 20	55 / 20	55 / 20	55 / 20	55 / 20
Raw comparison families	169	—	469	—	638	—
Synthetic comparison units	—	55	—	55	—	55
Eligible donor pool	—	159	—	462	—	621
Eligible donors per liable family	—	159	—	462	—	621
Observations	3,311	1,650	7,845	1,650	10,331	1,650
Regression units	224	110	524	110	693	110
Clusters	189	20	489	20	658	20
Unit fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Estimates Equation (28). September–November 2018 versus January–December 2019, excluding December 2018. All coefficients and standard errors are multiplied by 100. Revenue measures use fiscal-year 2018 accounts and are standardized at the mean and SD across the same 55 liable families. β is the liable-minus-comparison change at mean revenue shares; gradients are per liable-sample SD. All website-months receive equal weight. Actual DiD uses observed nonliable families; stacked SC includes one liable and one synthetic member per target-month. Nonnegative weights summing to one fit available June–November 2018 French histories with demeaning, donor-change noise scaling and ridge coefficient $\sqrt{12}$; there is no French-share penalty. Donors require complete French June–November 2018 and January–December 2019 observations, without foreign or global coverage requirements. Every target can use all 159, 462 or 621 donors in its selected pool. Actual regressions cluster by liable parent or control family; synthetic pairs share their liable-parent cluster (20 clusters). Synthetic intervals condition on weights and omit fitted-weight/shared-donor uncertainty. Overall and within R^2 and RMSE are reported separately for each specification. Advertising exposure requires a positive French-panel detection in September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The 638 nonliable families comprise 469 exposed and 169 with no observed qualifying exposure, including unknown exposure.

Table H.8: Fixed-weight sensitivity: Headline traffic estimates using advertising-exposed and pooled controls

Model Estimator Dependent variable	Advertising-exposed controls		Pooled controls	
	(1)	(2)	(3)	(4)
	DiD y_{it}	Stacked SC y_{ist}	DiD y_{it}	Stacked SC y_{ist}
Liabe $\times$ Post (β)	+7.52 (6.83)	+3.75 (6.29)	+7.06 (6.63)	+2.58 (6.21)
Liabe families	55	55	55	55
Liabe parent firms	20	20	20	20
Raw comparison families	469	–	638	–
Synthetic comparison units	–	55	–	55
Eligible donor pool	–	462	–	621
Eligible donors per liabe family	–	462	–	621
Regression units	524	110	693	110
Observations	7,845	1,650	10,331	1,650
Clusters	489	20	658	20
Unit fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes
R^2	0.911	0.994	0.914	0.994
Within R^2	0.00037	0.00175	0.00027	0.00084
RMSE (log units)	0.480	0.179	0.462	0.178

Notes: January–December 2019 versus September–November 2018; December 2018 excluded. Coefficients are 100β , with clustered standard errors in parentheses. Actual comparisons use family and month fixed effects; stacked synthetic comparisons use one actual liabe and one synthetic member per family-month, with member and shared month fixed effects. Synthetic weights fit available June–November 2018 French histories with demeaning, donor-change noise scaling, nonnegative unit-sum weights and ridge coefficient $\sqrt{12}$. Donors require complete French June–November 2018 and January–December 2019 observations, without foreign/global coverage restrictions; every target can use all 462 exposed or 621 pooled donors. The headline fitting procedure and validation are described in Section 4. Raw columns cluster by liabe parent or nonliabe family; synthetic pairs share the liabe-parent cluster. Inference conditions on fitted weights and omits fitted-weight/shared-donor uncertainty. Exposure requires a positive French-panel September–November 2018 detection on any family member domain of a study-designated liabe provider in the WTM advertising category. Adult advertising and custom overrides are excluded; unknown exposure remains in the residual group. Overall R^2 includes fixed effects; within R^2 absorbs them. Fit statistics do not establish counterfactual validity.

Table H.9: Fixed-weight sensitivity: Advertising-composition estimates using advertising-exposed and pooled controls

Estimator Model	Advertising-exposed controls		Pooled controls	
	Actual DiD (1)	Stacked SC (2)	Actual DiD (3)	Stacked SC (4)
Liabe $\times$ Post (β)	+7.52 (5.81)	+3.75 (4.78)	+7.06 (5.57)	+2.58 (4.77)
Liabe $\times$ Post $\times$ Composition (θ_C)	+12.38 (4.82)	+12.38 (4.96)	+12.38 (4.82)	+12.38 (4.96)
R^2	0.911	0.994	0.914	0.994
Within R^2	0.00147	0.03934	0.00117	0.03864
RMSE (log units)	0.480	0.175	0.462	0.175
Liabe families / parents	55 / 20	55 / 20	55 / 20	55 / 20
Raw comparison families	469	—	638	—
Synthetic comparison units	—	55	—	55
Observations	7,845	1,650	10,331	1,650
Regression units	524	110	693	110
Clusters	489	20	658	20
Unit fixed effects	Yes	Yes	Yes	Yes
Month fixed effects	Yes	Yes	Yes	Yes

Notes: Estimates Equation (12). September–November 2018 versus January–December 2019, excluding December 2018. All coefficients and standard errors are multiplied by 100. Revenue measures use fiscal-year 2018 accounts and are standardized at the mean and SD across the same 55 liable families. β is the liable-minus-comparison change at mean revenue shares; gradients are per liable-sample SD. All website-months receive equal weight. Actual DiD uses observed nonliable families; stacked SC includes one liable and one synthetic member per target-month. Nonnegative weights summing to one fit available June–November 2018 French histories with demeaning, donor-change noise scaling and ridge coefficient $\sqrt{12}$; there is no French-share penalty. Donors require complete French June–November 2018 and January–December 2019 observations, without foreign or global coverage requirements. Every target can use all 462 exposed or 621 pooled donors in its selected pool. Actual regressions cluster by liable parent or control family; synthetic pairs share their liable-parent cluster (20 clusters). Synthetic intervals condition on weights and omit fitted-weight/shared-donor uncertainty. Overall and within R^2 and RMSE are reported separately for each specification. Advertising exposure requires a positive French-panel detection in September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The 638 nonliable families comprise 469 exposed and 169 with no observed qualifying exposure, including unknown exposure.

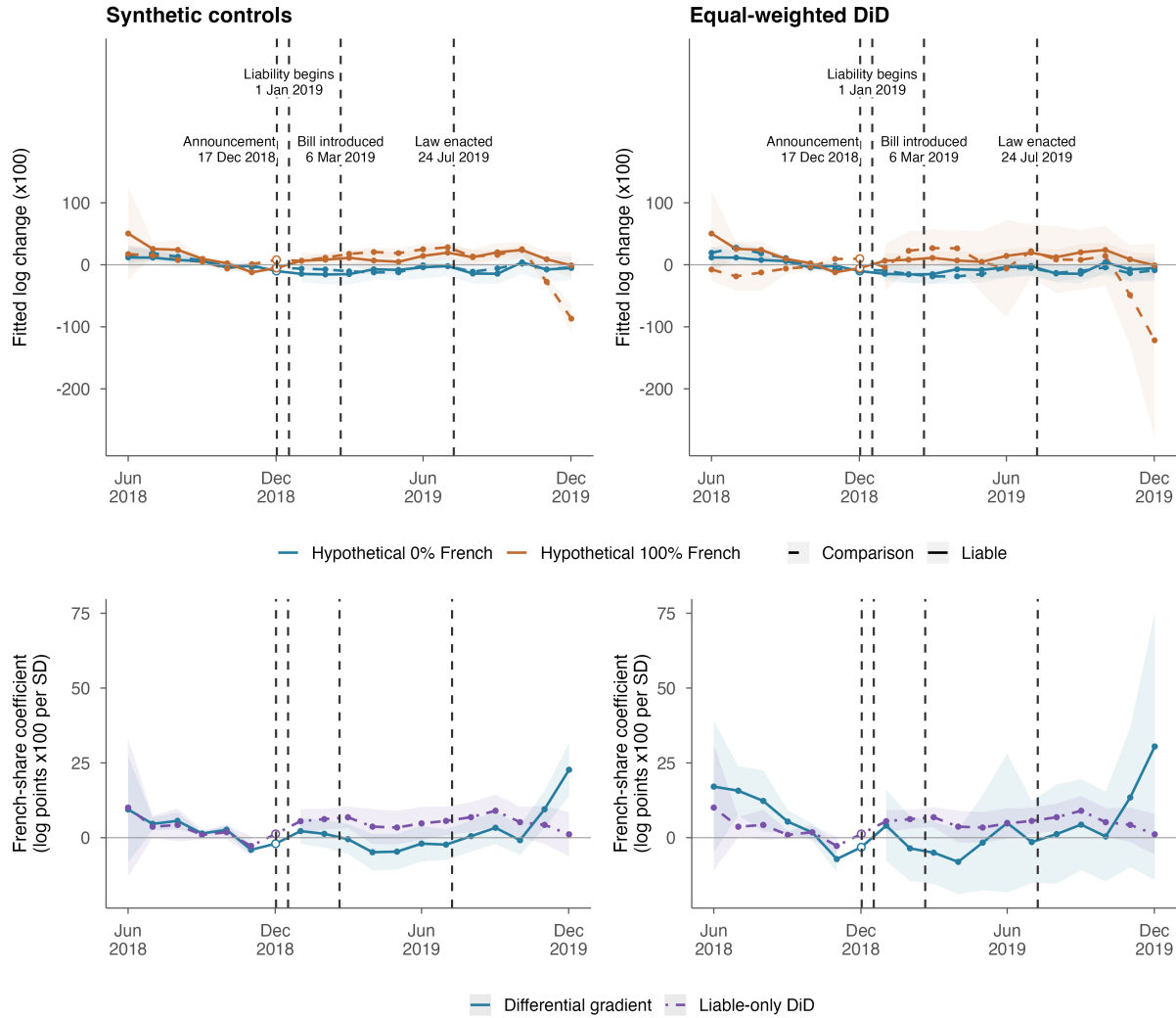


Figure H.4: Fixed-weight sensitivity: French-share heterogeneity with traffic-and-exposure-matched synthetic controls

Notes: Left: synthetic controls; right: equal-weighted actual-control DiD. Both use controls without documented advertising exposure and retain 55 liable families under 20 parents. Upper panels show fitted liable (solid) and comparison (dashed) paths at hypothetical 0% and 100% French traffic shares, using each column's own-member-share model and the original standardization. These endpoint profiles may extrapolate beyond observed support; they do not change the estimation sample. Lower panels show the differential gradient γ_m (blue; Table F.1, Panel A) and liable-share-only DiD gradient $\theta_{D,m}$ (purple; Panel B), omitting control-share interactions. Each purple curve uses its column's comparison method. Table F.1 retains advertising-exposed and pooled comparisons. Synthetic weights fit available June–November 2018 French traffic plus a share-gap penalty ($\kappa = 1$), with demeaning, donor-change noise scaling and ridge coefficient $\sqrt{12}$; balance is approximate. All 159 residual donors are eligible for every target, requiring complete French pre/post coverage without foreign/global restrictions. Shares use the Amazon.fr calibration conditional on 90% French traffic. Values are 100 times log coefficients; gradients are per original share SD (26.00 percentage points). Models jointly estimate June–November 2018 and January–December 2019, centered on September–November using full covariance. Hollow December points are auxiliary and display-only. Shading gives 95% clustered t intervals; synthetic intervals condition on weights and omit weight/shared-donor uncertainty. Monthly points lie at calendar midpoints; labeled lines mark announcement, liability, bill introduction and enactment. Exposure requires a positive French-panel detection during September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The residual group includes unknown exposure.

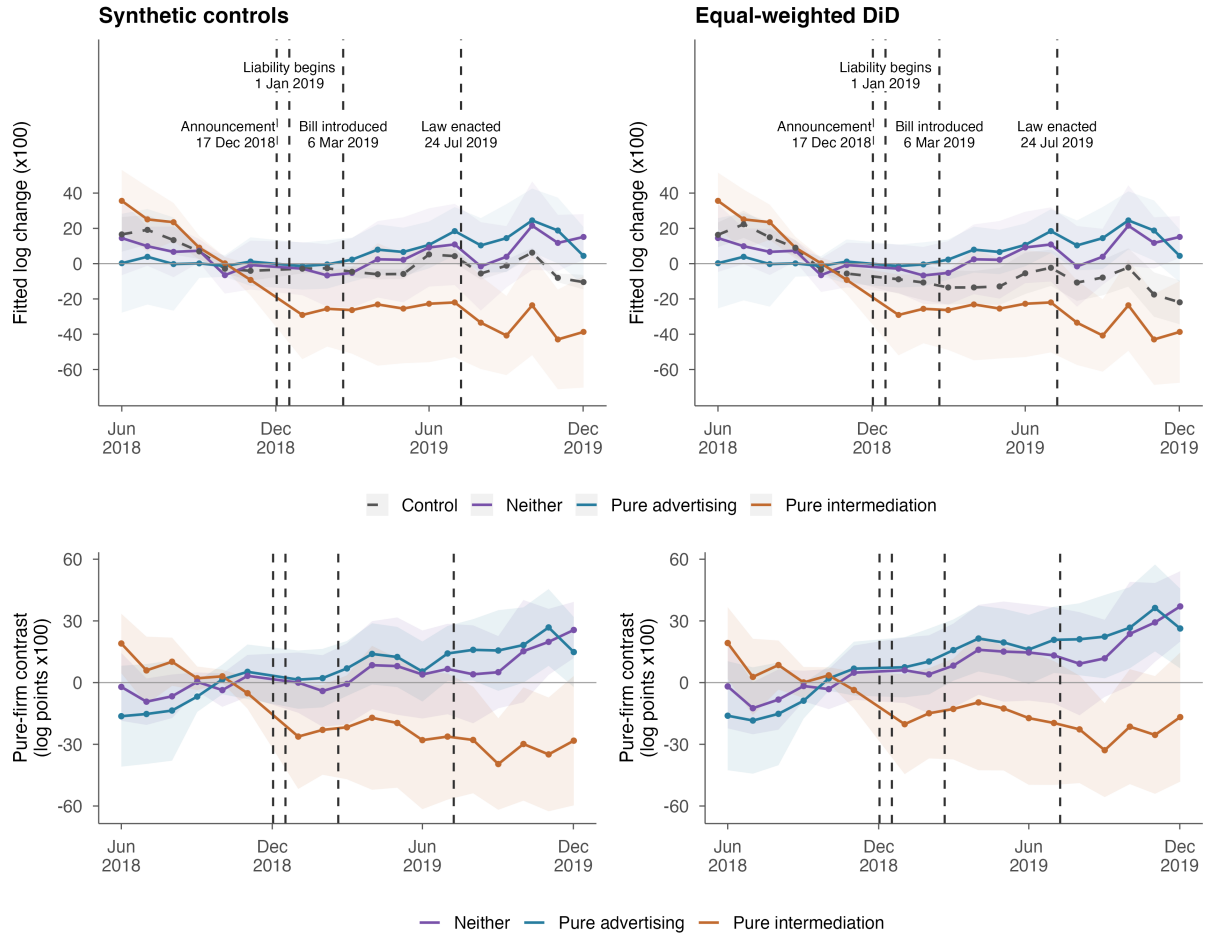


Figure H.5: Fixed-weight sensitivity: Advertising and intermediation revenue shares and traffic

Notes: Both columns use controls without observed advertising exposure and retain 55 liable families under 20 parents. Left: stacked synthetic-control DiD; right: equal-weighted actual-control DiD, corresponding to columns (2) and (1) of Table F.2. Upper panels show fitted comparison time effects (gray dashed) and hypothetical liable paths at $(A, I) = (1, 0)$ (pure advertising, blue), $(0, 1)$ (pure intermediation, orange), and $(0, 0)$ (neither activity, purple). These hypothetical profiles may extrapolate beyond observed shares; the neither profile is evaluated under liability despite zero in-scope revenue shares. Lower panels show liable-minus-control contrasts $\Delta_{A,m} = \beta_m + \theta_{A,m}(1 - \bar{A})/s_A - \theta_{I,m}\bar{I}/s_I$ and $\Delta_{I,m} = \beta_m - \theta_{A,m}\bar{A}/s_A + \theta_{I,m}(1 - \bar{I})/s_I$, and $\Delta_{N,m} = \beta_m - \theta_{A,m}\bar{A}/s_A - \theta_{I,m}\bar{I}/s_I$. All values are 100 times log coefficients, centered on September–November 2018 using full covariance. Models jointly estimate June–November 2018 and January–December 2019; December is excluded. Shading gives 95% clustered t intervals; synthetic intervals condition on fitted weights and omit weight/shared-donor uncertainty. Monthly points lie at calendar midpoints. Event lines mark 17 December 2018 (announcement), 1 January 2019 (liability), 6 March 2019 (bill) and 24 July 2019 (enactment). Synthetic weights use the headline French-only traffic criterion without French-share matching, with demeaning, donor-change noise scaling and ridge coefficient $\sqrt{12}$. All 159 residual donors are eligible for every target under complete French pre/post coverage without foreign/global restrictions. Exposure requires a positive French-panel detection during September–November 2018 on any family member domain of a study-designated liable provider in the WTM advertising category. Adult advertising and custom overrides are excluded. The residual group includes unknown exposure.