Political Effects of Newspaper Paywalls [*]

Julian Streyczek[**]

July 14, 2025

Abstract

I study how the introduction of paywalls on newspaper websites in the early 2010s affected political knowledge and electoral participation in the United States. Exploiting the staggered adoption of paywalls across newspapers, I first show that paywalls reduce readership of online news content, decreasing page views by a 25–30 percent on average. Next, I analyze the impact of county-level exposure to paywalls, measured as the share of online news consumption affected by paywalls, on political knowledge and electoral participation. I show that survey respondents in counties most exposed to paywalls become 1.6–2.4 percent less likely to correctly answer factual knowledge questions on party affiliations of political representatives and majorities in legislative bodies. This effect is driven by declines in knowledge about both national and regional politics, and is most pronounced among lower-income, less-educated individuals, and those least likely to subscribe to paid news. Additionally, higher paywall exposure reduces electoral participation among these non-subscribers by approximately 2 percent, particularly in federal and state congressional elections. These findings suggests that paywalls restrict access to politically relevant information, prompting substitution toward free alternatives with less political content. My results underscore the importance of easy access to high-quality news for democratic processes.

Keywords: Paywalls, News, Information, Politics

JEL: D72, L82, Z13

[*]I express deep gratitude to my advisors Carlo Schwarz, Guido Tabellini, and Luca Braghieri for their outstanding support. I also thank Alberto Manconi, Edoardo Teso, Elliott Ash, Eugen Dimant, Ho Kim, Jaime Marques Pereira, Jan Bakker, Jim Snyder, Julia Cagé, Nico Voigtländer, Rafael Jiménez-Durán, Ro'ee Levy, and Sarah Eichmeyer, as well as seminar and conference participants at Bocconi, Harvard, and more for valuable discussions and feedback. I gratefully acknowledge financial funding by Bocconi University, the Fondazione Invernizzi, and the German Academic Scholarship Foundation. All errors remain my own.

[**]Università Bocconi, Department of Economics. julian.streyczek@phd.unibocconi.it

1 Introduction

Over the past 15 years, paywalls have transformed much of newspapers' online content from a public good (back) into a club good: By the 2000s, most newspapers offered free online access to their articles, relying on advertising to generate revenue. However, as the shift to online readership undermined traditional print revenue streams, many publishers implemented paywalls on their websites – subscription models that require a monthly fee for full access – ending more than a decade of free news. By 2020, more than 80 percent of the most popular US newspapers had implemented such paywalls. Although these models help publishers remain profitable, they also restrict the availability of information, especially for those unwilling or unable to pay.

Access to information can shape voter knowledge and participation, which in turn can influence the accountability of elected representatives and the allocation of public funds (Besley and Burgess, 2002; Strömberg, 2004; Ferraz and Finan, 2008; Snyder Jr and Strömberg, 2010). Past shifts in digital media availability have prompted consumers to move away from news-intensive outlets, with significant consequences for voting patterns (Gentzkow, 2006; Drago et al., 2014; Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019).

In this paper, I study how the emergence of newspaper paywalls affected political knowledge and electoral participation, exploiting the staggered introduction of paywalls across time and geographic regions for causal identification. My analysis proceeds in two parts: First, I show that paywalls reduce consumption of newspapers' online content. Second, I examine the effect of paywalls on political knowledge and electoral participation by creating a measure of county-level exposure to paywalls and linking it to large-scale survey data.

In my analysis, I combine data on newspaper consumption, paywall introductions, and survey data on political knowledge and voting. I focus on the 100 largest daily newspapers in the United States based on print circulation in 2010, the year before the introduction of most paywalls, provided by the *Alliance for Audited Media*. For the largest 80 of these, I also measure online consumption using the daily number of page views between 2010 and 2017, collected by *Alexa Internet*. I assemble paywall introduction dates by checking mentions in news articles or historical website snapshots in the *Internet Archive*. Further, I measure individual-level knowledge and voting using the *Cooperative Election Study* [1], a yearly, nationally representative survey of the adult US population. I construct a political knowledge index using questions that ask participants to specify the party affiliation of

[1] Formerly called *Cooperative Congressional Election Study (CCES)*.

politicians and to name the majority party in US legislative bodies. I also use self-reported as well as validated data on participation in elections.

In the first part of my analysis, I identify the causal effect of paywalls on website page views using the sequential adoption of paywalls across websites. I compare newspapers that introduced a paywall between 2010 and 2017 to those that introduced paywalls later or never. I control for newspapers' market characteristics to account for trends among different audiences, and use estimators that are robust to heterogeneous treatment effects across newspapers or time. To corroborate the identifying assumption that trends in page views would be parallel in the absence of paywalls, I conduct two sets of robustness checks. First, I verify that selection of newspapers into paywalls is unlikely to drive the results, as pre-paywall trends in page views are parallel, inclusion of controls has minimal impact, and the result is robust to varying the composition of the control group. Second, I show that potential spillover biases, arising from readers' substitution of paywalled with non-paywalled outlets, are negligible, by conducting a series of robustness checks excluding newspapers that are expected to generate the highest spillovers.

I find that paywalls reduce the number of page views for paywalled newspapers by 25–30 percent, on average. These effects materialize in the first two months after the paywall and are persistent over the following years, suggesting that paywalls considerably displaced readers from mainstream US newspapers.

In the second part of my analysis, I start by quantifying the relevance of paywalls for survey respondents based on their county of residence. For this purpose, I construct a measure for counties' exposure to paywalls over time. This measure combines the timing of paywall introductions with newspapers' initial subscription rates in counties per household, which I validate as a strong predictor for regional online readership. Specifically, I define paywall exposure as the share of households subscribed to newspapers that have a paywall in place, which proxies the number of online readers of newspapers affected by paywalls. Then, I exploit variation in this measure using a staggered difference-in-differences design, classifying counties as treated once their paywall exposure exceeds the 75$^{\text{th}}$ percentile.[2]

Identification relies on the assumption that no other contemporaneous shock differentially influences political knowledge in treated counties at the time of paywall rollout, conditional on fixed effects and controls. I provide extensive evidence in support of this assumption. First, all specifications include state-by-year fixed effects that capture any time-varying shocks to the regional political landscape. Second, I include survey respon-

[2] The results hold when varying this threshold.

dent characteristics interacted with year fixed effects that capture changes in outcomes of certain demographics unrelated to paywalls, such as preferences for digital versus analog platforms. These variables also capture potential measurement error of paywall exposure correlated with regional characteristics. Third, the staggered nature of the treatment helps separate the effect of paywalls from aggregate trends.

I find that in counties with high paywall exposure, individuals are on average 1.2–1.8 percentage points (1.6–2.4 percent) *less* likely to answer political knowledge questions correctly, compared to a baseline accuracy of 74.5 percent correct answers. The event study rejects the hypothesis that prior trends in outcomes are driving the results. Interestingly, the declines in knowledge pertain to politicians and majorities on both state and federal level. Furthermore, heterogeneity analyses suggest that the effects are driven by the by groups most likely to be sensitive to price increases, namely lower-income and less-educated individuals.

As a placebo check, I show that these reductions in political knowledge can be explained by differences in the likelihood of purchasing news subscriptions. I use an additional survey to I predict subscription propensities based on individual characteristics. The effects of paywall exposure are entirely concentrated among those unlikely to pay for news, with no effect among individuals predicted to subscribe.

Finally, I investigate the effects of paywall exposure on participation in elections. Effects in the full sample are consistently negative but statistically insignificant. However, restricting the analysis to individuals predicted not to pay for news, and excluding highly partisan counties, yields larger and statistically significant reductions in participation. Both self-reported and validated turnout data indicate approximately a 2 percent decline in participation within this subgroup, driven primarily by lower turnout in federal and state congressional elections.

My results are most consistent with the interpretation that paywalls reduce news consumption from outlets providing extensive coverage of contemporary politics, particularly among individuals unwilling or unable to pay. These readers likely substitute toward free alternatives with less political focus, such as free news websites and social media, or disengage from news more broadly. As a result, they may pay less attention to politics, reducing political knowledge and ultimately electoral participation. While the absence of detailed individual-level consumption data prevents me from directly observing who reduces consumption of paywalled news and which substitutes they choose, several observations support this interpretation: First, substitution to newspapers' print versions is unlikely to explain the decline in page views, as aggregate print readership continued to decrease over the sample period, and case studies find little evidence for substitution to print (e.g., Pew

Research Center, 2023; Pattabhiramaiah et al., 2019). Second, the absence of paywall effects among likely newspaper subscribers contradicts the hypothesis that the effects are driven by changes to newspapers' content alongside paywall introductions. Instead, the result that declines in knowledge are concentrated entirely among likely non-subscribers supports the conclusion that switching away from traditional news sources drives the effect. Third, the finding that individuals with lower education and lower income are most affected by paywalls may reflect the monetary nature of the paywall barrier. Such individuals might have a lower willingness to purchase subscriptions due to relatively higher monetary costs.

My findings confirm that easy access to news is important for political knowledge and participation. Moreover, they underscore that in a predominantly profit-driven news market, economic shocks can ultimately affect democratic outcomes. To counteract these effects, modern democracies may benefit from guaranteeing broad and equitable access to relevant information.

This paper contributes to several streams of literature in economics and political science. First, it relates to existing work on news, knowledge, and accountability. Notable papers include the diffusion of television (Gentzkow, 2006; DellaVigna and Kaplan, 2007; Durante et al., 2019) and high-speed internet (Falck et al., 2014; Campante et al., 2018; Gavazza et al., 2019), which demonstrate that biased media can persuade voters, while substitution away from media with high news content can reduce political engagement. Studies on the entry and exit of newspapers underline that access to news can improve political participation (Gentzkow et al., 2011; Drago et al., 2014; Cagé, 2020; Gao et al., 2020; Djourelova et al., 2024), which may affect the performance of political representatives and the allocation of public resources (Besley and Burgess, 2002; Strömberg, 2004; Ferraz and Finan, 2008; Snyder Jr and Strömberg, 2010; Ash and Galletta, 2023).

To the best of my knowledge, I provide the first evidence on political effects of newspaper paywalls. Unlike most historical trends in media innovation and digitization, paywalls *increased* barriers to information. Moreover, I show that the *monetary* nature of these barriers disproportionally affected lower-income and less-educated individuals.

Second, my paper relates to the literature studying the news industry. Prior studies have shown that entry and exit of newspapers affect existing newspapers' revenues, prompting adjustments in topics, slant, or journalistic intensiveness (George and Waldfogel, 2006; Gentzkow et al., 2014; Angelucci and Cagé, 2019). Digitization has accelerated these pressures, as newspapers faced declining print advertising revenues and increased online competition (Seamans and Zhu, 2014; Bhuller et al., 2024). In response, many newspapers have introduced paywalls as part of their subscription models. Existing case studies indicate that paywalls may cause newspapers to lose readers (Cook and Attari, 2012; Chiou and Tucker, 2013; Pattabhiramaiah et al., 2019), although some outlets have succeeded in offsetting these losses through increased subscriptions (Chung et al., 2019; Aral and Dhillon, 2021). In a broader study, Kim et al. (2020) find that paywalls on 42 large US-based newspapers reduced page views by 30 percent on average.

I add to this literature in three ways. First, I provide the most comprehensive empirical study to date on how paywalls affect news consumption, studying the 100 largest US-based newspaper websites at the time. Second, I identify the causal effect of paywalls using state-of-the-art econometric methods. Third, I demonstrate that the effects on the news industry are both economically significant and persistent.

The remainder of this paper is organized as follows. In Section 2, I introduce my data and provide background on paywalls. In Section 3, I show how paywalls affected news consumption, and in Section 4, I analyze downstream effects on political knowledge and electoral participation.

2 Data

2.1 Newspapers

My analysis focuses on the 100 largest daily US newspapers by print circulation in 2010, based on data from the *Alliance for Audited Media (AAM)*. The dataset provides the number of physical copies sold for each newspaper per county in 2010, covering all counties where circulation exceeded 25 copies.[3] Exceptions are the three newspapers classified as national by *AAM*, namely the *New York Times*, *USA Today*, and *Wall Street Journal*, whose circulation is available at the Designated Market Area (DMA) level.[4]

To supplement information on online readership, I use website traffic data from *Alexa Internet*, a former web analytics service that was discontinued in 2021. It offered a free browser toolbar that provided website statistics to its millions of users, while collecting anonymized data on website visits. That scale makes *Alexa Internet* a uniquely valuable source of high-frequency readership data, providing reliable figures even for smaller local news websites. The dataset covers the daily number of page views from 2010 to 2017 for the 80 largest of the 100 newspapers in my sample. Page views reflect the total number of daily visits per million toolbar users, including all clicks to each newspaper's main website

[3]Specifically, circulation data for each newspaper correspond to a four-quarter period ending between Q1 and Q4 of 2011, representing the earliest year with complete coverage across newspapers in the sample.

[4]DMAs are 210 geographical regions in the United States, defined by the media analytics firm Nielsen.

and sub-domains.[5] Therefore, the variable should be interpreted as a proxy for total web traffic, which I verify using survey data. The data is limited to desktop users, thereby excluding traffic from mobile and tablet and does not provide information at the individual user level.

The newspapers in my sample represent a large share of total news consumption in the United States. They account for 76 percent of US print circulation recorded by *AAM* in 2010. According to *Alexa Internet*, their combined page views in 2010 were more than 1.5 times the combined page views of the websites of ABC, CBS, NBC, Fox, AP, and Reuters. Furthermore, none of the newspapers have exited the market as of the writing of this paper.

2.2 Paywalls

I obtain information on paywall launches through news coverage by the paywall-introducing newspapers themselves or competing outlets. Where ambiguous, I check historical website snapshots for the appearance of a digital subscription button using the Wayback Machine. Figure 1 shows the yearly share of newspapers in my sample that have a paywall on their website. Among the 100 newspapers, only five had introduced and upheld a paywall before 2010.[6] Between 2010 and 2016, 72 newspapers adopted paywalls, with an additional 14 doing so by 2021.

These paywalls were designed in a similar way: First, they usually adopted a ”metered” model, allowing free access to the full website up to a certain number of clicks per month (usually 10-20), after which a subscription was necessary to view additional content.[7] Among four newspapers, the paywall had the form of new, dedicated ”premium” website.[8] Second, paywalls were generally ”leaky”, meaning that they could be circumvented with sufficient effort (often intentionally), for example by clearing one’s browser cache or finding links to specific articles on search engines or social media. Therefore, paywalls increased either the monetary or non-monetary costs of access. For the purpose of this paper, I focus only on the existence of a paywall, abstracting from design features like the number of free articles, leakiness, and subscription pricing.

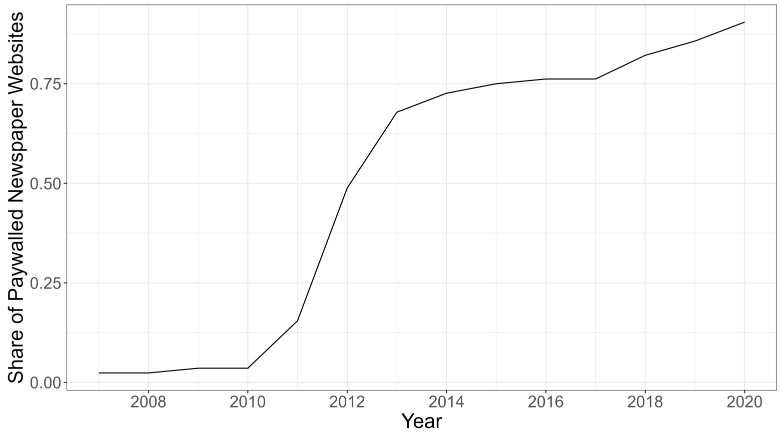
[5] *Alexa Internet* reported having several million toolbar users but never disclosed exact figures.

[6] Most notably, the *Wall Street Journal* introduced its paywall in 1997 and has maintained it continuously thereafter.

[7] The prominence of metered paywalls was partly due to Google’s ”First Click Free” policy, which until fall 2017 required news websites to grant each visitor at least three free articles per month, otherwise the website would be penalized in Google’s search ranking.

[8] The four newspapers are the *Boston Globe*, *Houston Chronicle*, *Philadelphia Inquirer*, and *San Francisco Chronicle*.

Figure 1. Share of newspapers with paywalls by year



Notes: Share of US newspapers that have ever implemented a paywall on their website, among largest 80 newspapers by number of page views in 2010.

2.3 Survey outcomes

I measure knowledge and voting in the US population using the Cooperative Election Study (CES) for the years 2006–2021 (Ansolabehere and Schaffner, 2022). The *CES* is an annual, nationally representative survey of the US population administered by *YouGov* that uses repeated cross sections of 10,000 to 60,000 respondents.

Political knowledge

I measure political knowledge using the following two types of questions:

1. *Which party has the majority in [legislative body]?*
2. *Which party does [name of political representative] represent?*

These questions are asked separately for four legislative bodies (US Senate, US House, State Senate, State House) and four political representatives (governor, House representative, two senators), based on the respondent's place of residence.

To avoid biases from multiple hypothesis testing, I compile responses to these eight questions into a political knowledge index by calculating the share of correct answers for each respondent in a given year. To account for possible correlations in the signal of the answers, I use as an alternative measure the first principal component of the eight answers, ensuring that a higher value corresponds to a larger share of correct answers.

Figure 2 presents summary statistics. On average, respondents answer 63 percent of questions correctly, where correct answers are most frequent for the governor, and least frequent for their state's Senate and House majorities. Across all questions, respondents may choose "I don't know", which I classify as an incorrect answer. A third of the respondents answers all questions correctly, indicating that my measure discriminates well between low and high, but less so between high and very high knowledge. The average knowledge index is relatively stable over time, with a slight upward trend. Note that the two more difficult questions on state legislative majorities were omitted in 2006 and 2009, producing higher values in these years.

Electoral participation

The *CES* includes self-reported data on voting in presidential, gubernatorial, and federal and state legislative elections, covering both House and Senate. These are collected in the weeks after elections, but have been shown to suffer from misreporting (Ansolabehere and Hersh, 2012). Therefore, I also use validated data that *YouGov* matches to official voting records based on survey respondents' identifiable information. Indeed, validated turnout is consistently 10–15 percentage points lower than self-reported participation. For comparison, I present both self-reported and validated measures throughout the analysis.

News consumption

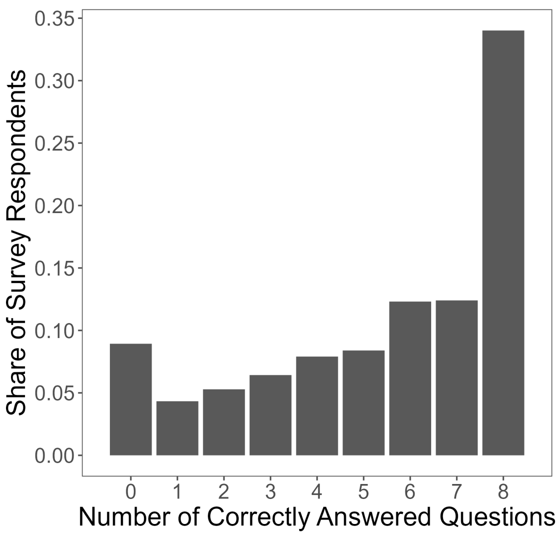
Data on news consumption is available only for selected years and measures whether individuals report consuming certain types of media in the 24 hours before answering the survey. In 2010, 56 percent reported reading a newspaper, of which 35 percent read exclusively online, 29 percent used both online and print sources, and 36 percent relied solely on print. Moreover, around 45 percent of respondents report that they follow the news "most of the time", the highest among four possible categories.

Paying for News

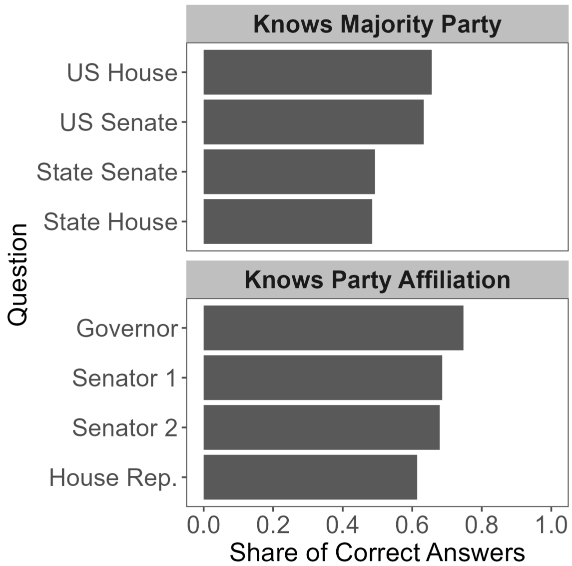
I use data from a nationally representative survey of US. adults conducted by the *Pew Research Center* and *Ipsos* (Pew Research Center, 2018), which includes 34,897 individuals surveyed between October 15 and November 8, 2018. It records whether individuals have paid for local news in the past year, along with detailed demographic information. According to the survey, approximately 14 percent of adults paid for local news sources in 2018. Although the data does not specify which local outlets respondents support and technically exclude national news subscriptions, they provide a rare and informative signal of the types of individuals generally willing to pay for newspaper content.

Figure 2. Descriptive statistics for political knowledge

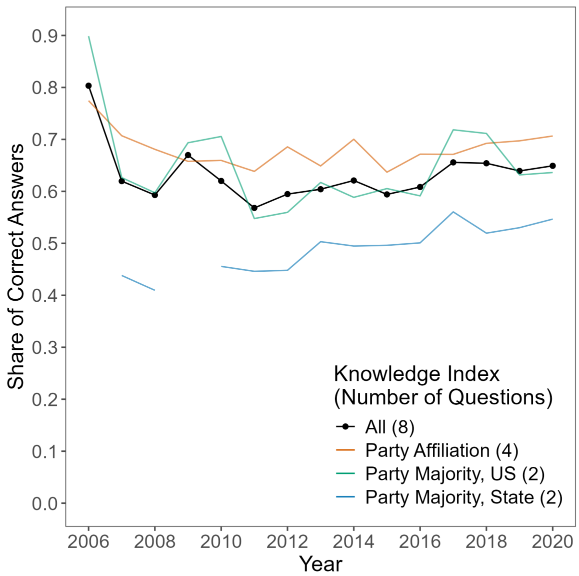
(a) Distribution of correctly answered political knowledge questions per respondent



(b) Share of correct answers by question



(c) Share of correct answers by year



Notes: The response "I don't know" is coded as an incorrect answer. The questions on State legislative majorities were not asked in years 2006 and 2009. Shares are computed among all questions that were answered by an individual (around 1 percent of individuals skip at least one answer). Survey weights are included. Panel (a) shows the distribution of the number of correct answers over survey respondents. Where less than eight questions were asked, I round the share of correctly answered questions, multiply by eight, and round to the closest integer. Panel (b) shows the number of correct answers by question. Panel (c) shows the share of correct answers among all questions asked in that year (black), as well as separately for the 4 questions regarding representatives' party affiliation, 2 questions regarding majorities in the US chambers, and 2 questions regarding majorities in the State chambers.

Table 1. Summary statistics for selected variables derived from CES

	Mean	SD	Min	Max	N
Political knowledge					
Knows majority: US Senate	0.63	0.48	0	1	520,495
Knows majority: US House	0.66	0.47	0	1	523,703
Knows majority: State Senate	0.49	0.50	0	1	475,198
Knows majority: State House	0.48	0.50	0	1	474,936
Knows party: Governor	0.75	0.43	0	1	528,258
Knows party: Representative	0.61	0.49	0	1	524,663
Knows party: Senator 1	0.69	0.46	0	1	527,706
Knows party: Senator 2	0.68	0.47	0	1	527,682
Knowledge index: Share correct answers	0.63	0.35	0	1	530,966
Vote participation (self-reported)					
Voted in general election	0.85	0.36	0	1	343,614
Voted in presidential election	0.71	0.45	0	1	212,796
Voted in gubernatorial election	0.66	0.47	0	1	196,897
Voted in House election	0.65	0.48	0	1	420,133
Voted in Senate election	0.68	0.47	0	1	282,406
Voted in State House election	0.64	0.48	0	1	379,062
Voted in State Senate election	0.63	0.48	0	1	373,816
Vote participation (validated)					
Voted in general election	0.54	0.50	0	1	420,712
Voted in presidential election	0.56	0.50	0	1	212,796
Voted in gubernatorial election	0.52	0.50	0	1	168,069
Voted in House election	0.51	0.50	0	1	383,760
Voted in Senate election	0.53	0.50	0	1	256,062
Individual Characteristics					
Age	46.95	17.00	18	109	556,746
Female	0.52	0.50	0	1	531,087
White	0.72	0.45	0	1	556,746
Has child	0.27	0.44	0	1	509,736
College degree	0.37	0.48	0	1	556,746
Post-graduate degree	0.10	0.30	0	1	556,746
Employed	0.46	0.50	0	1	556,746
Unemployed	0.22	0.41	0	1	556,746
Retired	0.19	0.39	0	1	546,796
Family Income $>$ 60k	0.35	0.48	0	1	556,746
Family Income $>$ 100k	0.15	0.36	0	1	556,746
Democrat	0.38	0.49	0	1	506,342
Republican	0.30	0.46	0	1	506,342
Media consumption					
Read newspaper in past 24h	0.48	0.50	0	1	384,282
Read newspaper online in past 24h	0.32	0.47	0	1	384,282
Follow news ”most of the time”	0.45	0.50	0	1	556,746

Notes: Selected summary statistics of survey respondent–level variables derived from CES for years 2006–2021, averaged across all years and including survey weights.

2.4 Other data sources

I incorporate data on county demographic and economic characteristics for 2010 from the 5-Year *American Community Survey* (U.S. Census Bureau, 2010). Moreover, I use county-level estimates for visits to selected newspaper websites from the *MRI-Simmons Local* dataset, which assigns responses to its *National Consumer Survey* (MRI-Simmons, 2014) to geographies based on respondent and county characteristics. I measure county partisanship using the partisanship index in the *National Neighborhood Data Archive (NaNDA)* (Chenoweth et al., 2020), which is based on Democratic and Republican vote shares in past elections. To link counties with media markets, I use the crosswalk from Gentzkow and Shapiro (2008).

3 Effect on news consumption

In this section, I analyze how the launch of paywalls on major US newspaper websites between 2010 and 2017 affected page views. I employ a staggered difference-in-differences design that exploits the sequential timing of paywall introductions across newspapers.

3.1 Empirical strategy

My analysis is based on the following two-way fixed effects (TWFE) regression model:

$$\log(Pageviews_{\,n,t}) = \sum_{\tau \neq -1} \beta_\tau \, D^{\tau}_{n,t} + \alpha_n + \gamma_t + \delta' \, \mathbf{X_{n,t}} + \varepsilon_{n,t} \tag{1}$$

$Pageviews_{\,n,t}$ denotes the number of page views for newspaper n in year-month t. I apply a log transformation to account for skewness, enabling interpretation of the coefficients as (approximate) percentage effects on page views. $D^{\tau}_{n,t}$ is an indicator equal to one if and only if newspaper n introduces a paywall τ months after t. α_n and γ_t denote newspaper and year-month fixed effects, respectively. I cluster standard errors at the newspaper level.

To account for differential trends in newspapers' page views as well as shocks to their respective audiences over time, I include in $\mathbf{X_{n,t}}$ characteristics of the markets in which the newspapers operated as of 2010, interacted with year-month fixed effects. Specifically, using the counties in which the newspaper is circulated, I compute the average population density, the shares of population with yearly income below $\$50k$ and above $\$100k$, the share of college-educated individuals, and the *NaNDA* partisanship index, weighted by the number of print copies sold in each county.

For this analysis, I restrict my sample as follows: First, I exclude the four newspapers that introduced new premium websites as part of their paywalls. Second, I exclude the newspaper *Newsday*, which introduced a paywall in October 2009, just before the sample period. Third, I exclude three small newspapers with significant data gaps, likely due to data collection errors by *Alexa Internet*. The final sample comprises the monthly page views for 72 newspapers from 2010 to 2017. Among these newspapers, two introduced paywalls before the sample period, 55 during the sample period, eleven afterward, and four never introduced a paywall (as of 2022). In my preferred specification, I assign the two always-treated and eleven later-treated newspapers as untreated, in addition to the four never-treated. These newspapers arguably act as valid control units because they implemented their paywalls well outside the sample period. I demonstrate that my results are robust to varying definitions of the control group.

3.2 Identification

Identification of the monthly treatment effects β_τ relies on the assumption that page views of paywalled newspaper websites would have evolved parallel to those of non-paywalled websites in the absence of paywalls. In this sub-section, I discuss several challenges to this assumption.

Heterogeneous effects

I account for potential biases from heterogeneous treatment effects across units or time, which may arise when estimating Equation (1) via Ordinary Least Squares (for example, see Baker et al., 2022). I use the robust estimator proposed by Callaway and Sant'Anna (2021), which aggregates doubly-robust estimates for treatment group-time effects using only "clean" control units (Sant'Anna and Zhao, 2020).[9] Control variables are included in each of those group-time-specific estimates both as a regression adjustment and for the construction of probability weights. Additionally, I employ the robust estimator by Sun and Abraham (2021), which estimates treatment effects by including only treated and never-treated units in the outcome regression and re-weighting the estimates accordingly.

Selection

Since the decision to introduce a paywall is endogenous, paywalls may correlate with unobserved factors not accounted for in the regression. To mitigate concerns about selec-

[9]Throughout the analysis, I use a common ("universal") base period, consistent with Equation (1), estimating newspaper-specific treatment effects relative to the month preceding paywall introduction.

tion, I do the following. First, I show that pre-treatment trends in page views are parallel, ensuring that such trends are not biasing the result. Second, I control for market-based factors interacted with month-year fixed effects, which accounts both for different incentives for newspapers to introduce a paywall, and for time-varying shocks to different types of newspaper audiences. Third, to address concerns about selection in the control group, I show that results are statistically and economically significant for different compositions of the control group. Fourth, I explore heterogeneity in treatment effects to ensure that the results are not just driven by a small subset of newspapers.

Spillovers

Another potential concern involves biases from spillover effects: Readers displaced by a paywall may switch to a newspaper in the control group, which leads to an upward bias (in absolute terms) in the estimate for the paywall effect. Formally, such spillovers violate the stable unit treatment value assumption (SUTVA), which requires that units' potential outcomes do not depend on the treatment status of other units. However, in this setting, spillovers are unlikely to be a first-order concern due to the segmented nature of the newspaper market: All but three newspapers are classified as regional by *AAM*, catering primarily to their respective regional markets. As a result, these newspapers likely act as poor substitutes for one another. For example, it seems unlikely that a reader of the *Chicago Tribune* would respond to a paywall by switching to the *Los Angeles Times*.

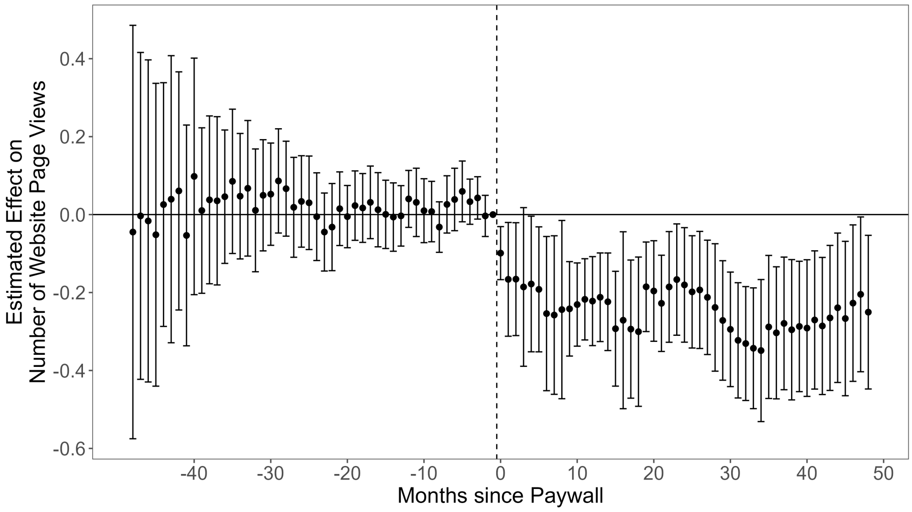
To address any remaining concerns, I verify that my results hold when removing the largest paywalled newspapers and the smallest control newspapers. Intuitively, paywalls on larger newspapers displace more readers, whose substitution has a larger relative effect on the control newspapers the fewer readers these newspapers have.

3.3 Results

Main result

Figure 3 presents estimates for the dynamic effect of paywalls on website page views, using the robust Callaway and Sant'Anna (2021) estimator. On average, this "first wave" of paywalls led to a sharp and persistent reduction in the number of page views for paywalled newspapers. Table 2 reports estimates for the static effect. On average, page views decreased by 25–30 percent, depending on the choice of control variables. Figure A2 and Table A1 demonstrate that these results are robust to using alternative estimators.

Figure 3. Effect of paywalls on website page views



Notes: Monthly coefficients for β_τ from Equation (1): Regression of monthly number of newspaper website views on indicators denoting the number of months since paywall implementation. The omitted category is the month before the paywall. Includes newspaper fixed effects, year-month fixed effects, and log population density, shares of three income buckets, college-educated share, and partisanship index for newspapers' audiences, interacted with year-month fixed effects. Estimated using Callaway-Sant'Anna estimator. Vertical bars denote 95 percent pointwise confidence intervals. Standard errors are clustered by newspaper.

Table 2. Effect of paywalls on pageviews

Dependent Variable:	Log(Pageviews)				
	(1)	(2)	(3)	(4)	(5)
Paywall	-0.257***	-0.291***	-0.302***	-0.280***	-0.310***
	(0.070)	(0.062)	(0.064)	(0.063)	(0.076)
Newspaper FE	✓	✓	✓	✓	✓
Month-Year FE	✓	✓	✓	✓	✓
Controls x Month-Year FE:					
Log(Population Density)		✓	✓	✓	✓
Sh. HH income $\leq 50k$			✓	✓	✓
Sh. HH income $> 100k$			✓	✓	✓
Partisanship index				✓	✓
Sh. college-educated					✓
Observations	6,912	6,912	6,912	6,912	6,912

Notes: Coefficients for static (pre-post) version of Equation (1): Regressions of monthly number of newspaper website views on indicator denoting active paywall. All specifications include newspaper fixed effects and year-month fixed effects. Controls represent average county characteristics at 2010-levels where newspaper is active, weighted by number of print copies sold, interacted with year-month fixed effects: Log population density, share of households with yearly income below $50,000 and above $100,000, share adults with college degree, and NaNDA partisanship index. Estimated using Callaway-Sant'Anna estimator. Standard errors in parentheses are clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Interpretation

The dynamic estimates reject the presence of pre-treatment trends, indicating that newspapers did not select into introducing paywalls based on prior trends in page views. The immediate decline after introducing a paywall supports the interpretation that paywalls raise the opportunity cost of website access, prompting some users to reduce their visits. Conversely, paywalls may also encourage new subscribers to visit more frequently. Thus, the estimates represent the net outcome of these two opposing effects. However, without individual-level user data, it is not possible to determine whether the observed decline reflects the displacement of a few highly active readers or many rather casual users.

The timing of the effects is inconsistent with alternative explanations that rely on gradual changes accompanying paywalls, such as shifts in topic composition or ideological slant, as these would likely materialize more slowly. Nonetheless, I cannot entirely rule out that these factors contributed to the modest downward trend observed in the treatment effects. However, other long-term factors may also explain this trend. For instance, the diffusion of broadband and 3G internet during the sample period likely increased overall interest in online compared to offline news. If new users entering the online news market disproportionately favored non-paywalled over paywalled websites, the initial effect of paywalls would become more pronounced over time.

How many individuals respond to paywalls by switching to the print version of the paywalled newspaper? While I cannot provide a definite answer based on the available data, this channel is unlikely to explain the observed effects. During the sample period, digital news consumption grew rapidly while print news consumption declined: The growth in the number of newspaper website visits often exceeded 10 percent *per year*, while the number of print copies sold declined by around 30 percent between 2009 and 2017 (Pew Research Center, 2023). Moreover, studying the *New York Times* paywall, Pattabhiramaiah et al. (2019) conclude that spillovers from digital to print were negligible, with a 17 percent decline in website visits accompanied by only a 1-4 percent increase in print circulation. This increase was primarily driven by bundling online subscriptions with the print version, rather than by substitution to standalone print subscriptions.

Further robustness

In this sub-section, I confirm that my results are robust to changes in the choice of which newspapers constitute the control group, and are not driven by users' substitution between paywalled and non-paywalled newspapers.

Table 3 shows results under different control group compositions. Column (2) shows that excluding the pre-treatment period observations for the 55 not-yet-treated newspapers from the control group has minimal impact on the results. Columns (3) and (4) adopt the most restrictive definition for the control group, comparing newspapers that implemented a paywall during the sample period to those that introduced one just afterwards. If anything, the effects are larger under these specifications, with a 34–36 percent decline in page views compared to the baseline estimate of a 28 percent decrease.

Table A2 shows results when excluding different sets of newspapers from the sample. Columns (1) and (2) indicate that excluding large newspapers, in particular the three national newspapers and the fifteen largest newspapers by page views, increases the estimate slightly. Columns (3) and (4) exclude the fifteen largest paywalled newspapers and the five smallest non-paywalled newspapers. If spillovers were significant, one would expect the absolute magnitude of the estimates to decrease under these restrictions. However the opposite is true: the estimates are either unchanged or increase slightly. Intuitively, while theoretically an issue, spillovers may not play a significant role empirically because of the regional segmentation of the news markets, making most newspapers poor substitutes for one another. Moreover, the control group is sufficiently large to dilute potential biases.

Table 3. Effect of paywalls on pageviews - Control groups

Dependent Variable:	Log(Pageviews)			
Control group:	All	w/o Treated	w/o Earlier/Later	Only Later
	(1)	(2)	(3)	(4)
Paywall	-0.280***	-0.277***	-0.342***	-0.355***
	(0.067)	(0.070)	(0.068)	(0.068)
Newspaper FE	✓	✓	✓	✓
Month-Year FE	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓
Control group:				
Earlier Treated	✓	✓		
Treated (pre-treatment)	✓		✓	
Later Treated	✓	✓	✓	✓
Never Treated	✓	✓		
Observations	6,912	6,912	6,336	6,336

Notes: Coefficients for static (pre-post) version of Equation (1), for different compositions of the control group: ”Earlier treated” are 2 newspapers that implemented paywall before 2010. ”Treated (pre-treatment)” are 55 newspapers that implemented paywall between 2010 and 2017. ”Later Treated” are 11 newspapers that implemented paywall after 2017. ”Never Treated” are 4 newspapers that have not implemented a paywall by 2022. Regressions of monthly number of newspaper website views on indicator denoting active paywall. All specifications include newspaper fixed effects and year-month fixed effects. Controls represent average county characteristics at 2010-levels where newspaper is active, weighted by number of print copies sold, interacted with year-month fixed effects: Log population density, share of households with yearly income below \$50,000 and above \$100,000, share adults with college degree, and NaNDA partisanship index. Estimated using Callaway-Sant'Anna estimator. Standard errors in parentheses are clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4 Effect on political knowledge and electoral participation

In this section I explore how paywalls affected factual knowledge about contemporary US politics as well as participation in elections by leveraging temporal and regional variation in exposure to paywalls.

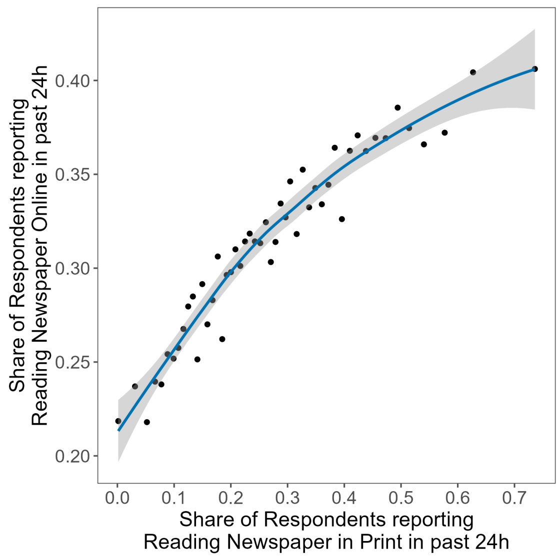
4.1 Measuring regional exposure to paywalls

A key challenge is the lack of granular data that directly links web traffic from specific counties to individual websites.[10] As a practical alternative, I proxy regional online readership using newspapers' regional print circulation, which provides high-quality information on county level. For regional newspapers, which constitute the majority of my sample, print circulation reveals in which localities the regional focus of the newspaper is relevant. For the *New York Times*, instead, print circulation rather captures where its predominantly left-leaning and highly educated audience resides.

To validate this approach, I present two pieces of evidence. First, Figure 4 presents a binned scatter plot for the shares of respondents who report reading a newspaper offline versus online in the past 24 hours, averaged by county and year. It reveals that consumption of newspapers offline is strongly positively correlated with consumption online, suggesting that print circulation successfully captures regional variation in the overall intensity of engagement with newspaper content, regardless of format. Second, to verify that this relationship also holds for individual newspapers, I use *MRI-Simmons* data on estimated website visits by county for the *Washington Post* and *Los Angeles Times* – the only two non-national newspapers for which this data is available. Figures 5a and 5b show scatter plots for the number of print copies sold per household in a given county, versus the (estimated) share of the population visiting the respective website in the past 30 days. In both cases, higher print circulation is positively and approximately linearly associated with more frequent website visits for these websites. Together, these results suggest that newspapers' print circulation captures relevant variation in the regional engagement with their respective websites.

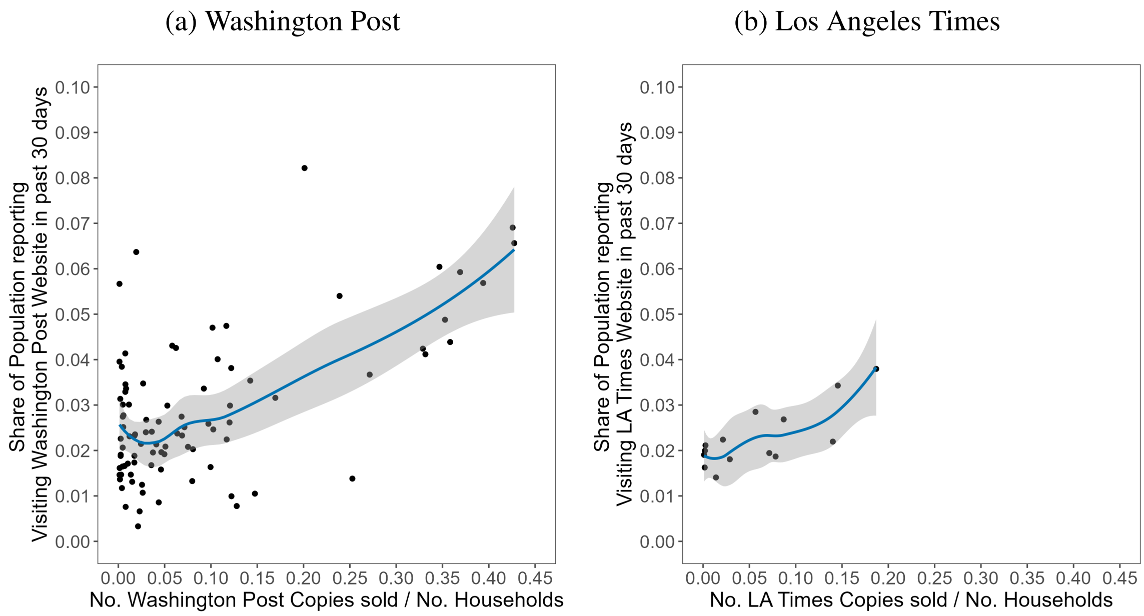
[10]For example, at the beginning of my sample, *comScore*'s browser panel of 50,000 individuals records fewer than 100 unique weekly visitors to most newspaper websites outside the top three.

Figure 4. County-level correlation of newspaper readership online vs. offline



Notes: Binned scatterplot using self-reported newspaper readership data from the CES 2006–2021. The x- and y-axes show the share of individuals who report reading a newspaper in print and online, respectively, in the past 24 hours. Observations are aggregated at the county-year level prior to binning. Sample weights are included. The blue line displays the LOESS-smoothed conditional mean with 95 percent confidence intervals.

Figure 5. County-level correlation of print circulation and website visits for selected newspapers



Notes: Each dot represents a county. Panel (a) shows data for the Washington Post; Panel (b) for the Los Angeles Times. The y-axis reports the percent of respondents who visited the newspaper's website in the past 30 days in 2013, based on county-level MRI-Simmons estimates constructed via multilevel regression and post-stratification (MRP). The x-axis shows the number of print copies sold per household in 2011, including counties with print circulation greater than 100 and excluding outliers at the top and bottom of both axes. The blue line displays the LOESS-smoothed conditional mean with 95 percent confidence intervals.

Motivated by this evidence, I define paywall exposure for each county and year as the share of households subscribed to newspapers that have a paywall in place. Formally:

$$Paywall\ Exposure_{c,t} = \sum_{n \in Newspapers} \frac{Circulation^{\,n}_{c,2010} \times I(Paywall^{\,n}_{t})}{Households_{c,2010}} \tag{2}$$

where $Circulation^{\,n}_{c,2010}$ is the number of print copies sold of newspaper n in county c in 2010. $I(Paywall^{\,n}_{t})$ is a binary variable that indicates whether newspaper n has implemented a paywall by year t. I fix circulation at their 2010 levels, prior to the widespread introduction of paywalls, to avoid endogeneity arising from paywalls themselves potentially affecting circulation. This measure of paywall exposure resembles shift-share variables (see Bartik, 1991; Blanchard et al., 1992), where initial market shares serve as shares and paywall introductions act as shifts.

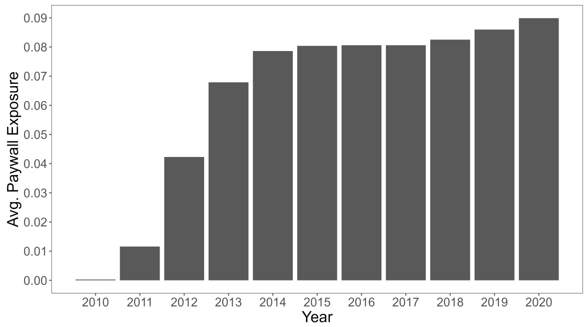
To provide insights into the variation captured by paywall exposure, Figure 6 presents key descriptive statistics. Panel (a) shows that paywall exposure increased significantly between 2010 and 2014, and plateauing afterward. Panel (b) highlights substantial variation across counties, with only few experiencing the highest levels of exposure. To account for differences in county populations, Panel (c) displays the distribution of paywall exposure among respondents in the *CES* survey. Accounting for county sizes reveals that several large counties experience intermediate to high paywall exposure, producing a less skewed distribution. Note that since paywall exposure is intended as a proxy variable, its relative differences across counties and years are more informative than exact numerical values.[11]

Figure 7 shows geographic variation, revealing that paywalls affected all major regions of the US, while concentrating around urban areas. To gain a deeper understanding of which regions are most likely to be affected by paywalls, Figure A4 presents coefficients from a regression of paywall exposure in 2020 on (standardized) county characteristics in 2010. Paywalls disproportionately impacted high-income, Democratic-leaning counties. However, paywalls do not appear to be correlated with political knowledge, conditional on controls.

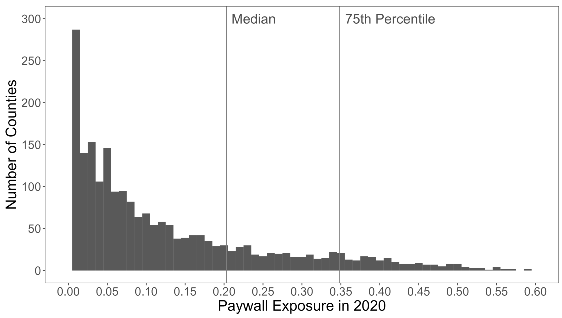
[11] In particular, paywall exposure should not be interpreted as the number of households affected by paywalls, for the following reasons: First, a household may subscribe to multiple newspapers. Second, each household member may have access to multiple distinct subscriptions for the same newspaper (e.g., through their employer or university). Third, because print and digital newspaper access are often bundled, subscribing households are typically *not* restricted by paywalls.

Figure 6. Paywall Exposure

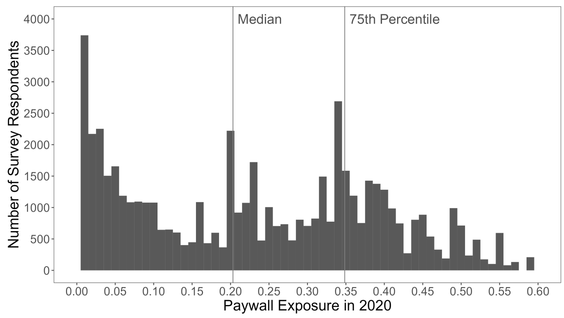
(a) By year



(b) Distribution over counties

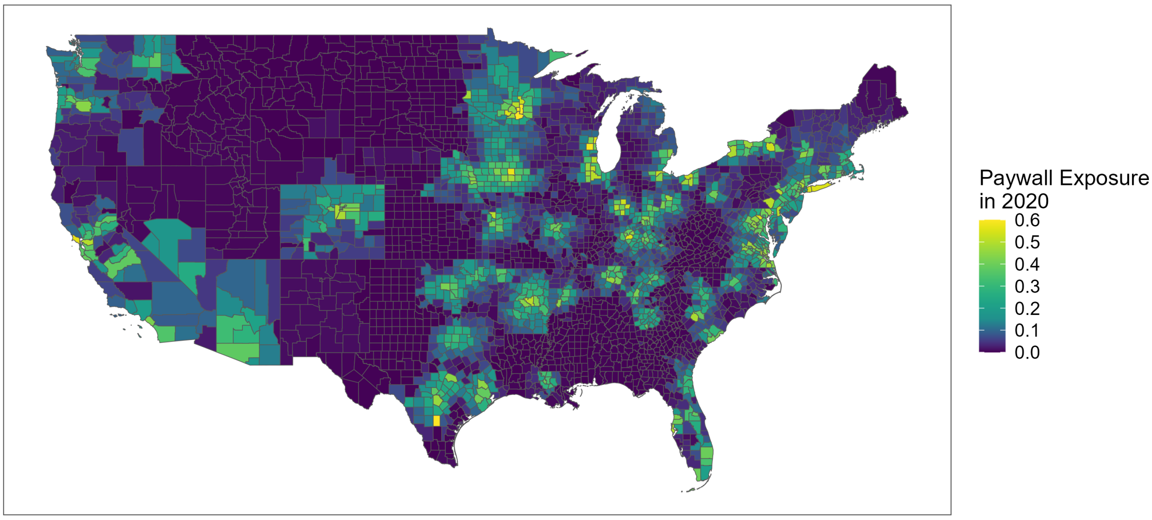


(c) Distribution over survey respondents



Notes: Panels (a) and (b): Average paywall exposure across counties by year. Panels (c) and (d): Distribution of paywall exposure over counties and survey respondents (weighted), respectively, in 2020.

Figure 7. Geographical variation of paywall exposure



Notes: Paywall exposure by county in 2020.

4.2 Empirical strategy

I exploit the staggered exposure of counties to paywalls using a difference-in-differences design. For this purpose, I divide counties into a low- and a high-paywall exposure group, with counties entering treatment once paywall exposure exceeds a fixed threshold (Guriev et al., 2021). The baseline specification is based on the 75th percentile of paywall exposure at the end of the sample period, which ensures high power by splitting observations roughly equally into treated and untreated observations, across all years. In robustness checks, I confirm that the results are robust to varying this threshold.

The causal analysis is motivated by the following specification:

$$y_{\,i,c,t} = \sum_{\tau \neq -1} \beta_\tau \, D^{\tau}_{c,t} + \delta' \mathbf{X}_{i,c,t} + CountyFE_c + YearFE_t + \varepsilon_{i,c,t} \tag{3}$$

The dependent variable $y_{\,i,c,t}$ is an outcome variable for survey respondent i residing in county c in year t, such as the share of correctly answered political knowledge questions. County fixed effects account for pre-existing regional differences in knowledge, for example due to education or political composition, and year fixed effects capture yearly shocks to knowledge. Alternative specifications include state-by-year fixed effects instead, which capture changes in knowledge due to elections.[12]

$\mathbf{X}_{i,c,t}$ denotes time-varying control variables. In the baseline specification, these include individual demographic variables in the form of categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2). In robustness checks, I interact these with year fixed effects. I cluster standard errors at the county level.

The treatment variables $D^{\tau}_{c,t}$ are indicators equal to one if and only if paywall exposure in county c first exceeds the sample-weighed 75th percentile τ years after t. Splitting at the 75th percentile (as opposed to the median, for example), ensures a suitable number of treated and control observations across the panel. In particular, roughly equal parts of observations are never-treated, treated early (by 2017), and treated late (by 2021). Therefore, this approach ensures high power.

I estimate Equation (3) using the robust estimator proposed by Callaway and Sant'Anna (2021).[13] In each year, the control group consists of counties that are either never treated

[12] Note that the specification does not include individual fixed effects because the survey is a repeated cross section in which each individual appears only once.

[13] Throughout the analysis, I use a common ("universal") base period, consistent with Equation (3), estimating county-specific treatment effects relative to the year preceding treatment.

or not yet treated. While county fixed effects are implemented natively, this is not the case for state-by-year fixed effects and individual-level controls.[14] Therefore, I proceed in two steps. First, I residualize the outcome variable on the individual controls, county fixed effects, and state-by-year fixed effects. Second, I apply the Callaway and Sant'Anna (2021) estimator to these residualized outcomes. I verify empirically that the inclusion of control variables in this way improves precision of the estimator without considerably changing the point estimates.

4.3 Identification

I identify the average effect of paywall exposure on political knowledge across the adult US population, comparing regions with high versus low exposure. Identification requires that, in the absence of paywalls, trends in political knowledge would have evolved similarly in high- and low-exposure regions. In the following, I discuss four key challenges to identification and strategies to address them: Omitted county characteristics, correlations with regional politics, heterogeneous effects, and selection of newspapers.

Omitted county characteristics

A primary concern is that paywall exposure may be correlated with county-level characteristics that independently affect political knowledge. For example, in Section 4.1 I show that paywall exposure is higher in counties with higher initial income. If such counties experienced shocks to political knowledge that are unrelated to paywalls, estimates may be biased. For instance, high-income individuals may have a higher propensity to respond to the rise of social media by substituting away from newspapers toward softer news and entertainment content, even in the absence of paywalls. If not accounted for, this could lead to a downward bias in the estimated effect of paywalls.

More subtly, this bias may also arise from non-random measurement error in paywall exposure. If high-income counties have higher subscription shares for a given level of online engagement, then paywall exposure is overstated in these areas. Conversely, if paywalls are on average more restrictive in high-income counties, actual exposure may be understated. In both cases, failing to account for potential shocks to high-income individuals may bias estimates.

[14]Although the estimator allows for the inclusion of control variables, their implementation is considerably more demanding than specified in Equation (3): Controls enter the estimation of each group-time average treatment effect, effectively interacting them with treatment-group-by-year fixed effects. This approach requires sufficient support in each cell, and estimation often fails when group sizes are small.

I address these concerns in three main ways. First, I verify that results are robust to including interactions of individual-level controls with year fixed effects. These coefficients flexibly absorb year-specific shocks to specific subpopulations, such as changes in media consumption habits. I also check robustness to including state-by-year fixed effects, which account for unobserved state-level shocks. Second, the staggered implementation of paywalls helps distinguish their effects from contemporaneous trends. Even if some regions exhibit changes in political knowledge due to unobserved factors, it is unlikely that the timing of these trends would systematically coincide with the timing of paywall adoption across counties. Third, I show that results are robust to alternative threshold definitions for discretizing counties into high- and low-exposure groups. This reduces sensitivity to measurement error near the cutoffs and ensures that results are not driven by misclassification of a small number of units.

Correlation with regional politics

The identifying assumption may also be violated if regions with high paywall exposure experienced divergent trends in political knowledge for reasons unrelated to paywalls. For instance, if states with more paywalls happen to elect many newcomers shortly before paywall adoption, residents might find political survey questions easier or harder to answer, leading to lower measured knowledge.

To address this concern, I estimate two sets of specifications, one with only year fixed effects, and one that instead includes state-by-year fixed effects. The former leverages variation in paywall exposure across all counties and years, and should therefore expected to have more power. The latter relies on comparisons of counties within state and year, effectively holding constant statewide political shifts such as election outcomes, political events, or policy changes.

Heterogeneous Effects

Survey respondents may respond differently to paywalls depending on individual characteristics. For example, initial news consumption patterns may play a role. Holding paywall exposure fixed, high-income counties may receive a smaller share of news from television, making them more affected by paywalls. In this case, one would expect larger effects in high-income areas. Alternatively, high-income individuals may be more likely to purchase subscriptions in response to paywalls, while low-income individuals may switch to free news sources with less news content, implying smaller effects in high-income areas. Moreover, the effect of paywalls may also depend on their business and editorial decisions as well as the presence of other local newspapers, and may unfold or vary over time.

Such heterogeneous effects do not pose a threat to identification but are important for interpretation. The Callaway-Sant'Anna estimator accounts for heterogeneity by ensuring that comparisons are made between appropriate units and that group-time specific treatment effects are aggregated correctly. Therefore, it explicitly allows that different groups experience effects with varying intensity and and different patterns over time. However, when interpreting the results, it is important to keep in mind that the reported average effect may mask substantial heterogeneity across individuals and counties.

Selection of Newspapers

Finally, one may be concerned that omitting newspapers outside the largest 100 in the analysis could bias the constructed measure of paywall exposure. In fact, this course of action effectively ignores any paywalls of these smaller newspapers, leading to underestimation of paywall exposure in these regions.

To address this concern, I verify that the results are robust to excluding counties where omitted newspapers account for a large share of total circulation. I also check robustness to dropping small counties, as these are more likely to rely on smaller, local newspapers outside the top 100.[15]

4.4 Effects on political knowledge

In this subsection, I present results for the effects of paywall exposure on political knowledge.

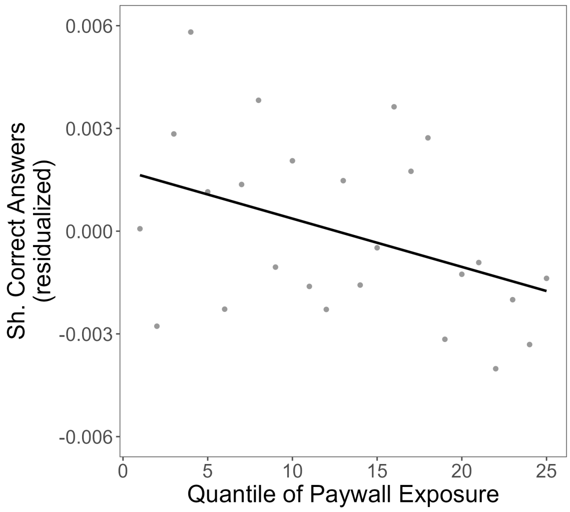
Motivating evidence

I start by providing motivating evidence on the negative correlation between paywall exposure and political knowledge. Figure 8 plots 25 quantiles of paywall exposure against the share of correctly answered political knowledge questions (residualized on county, year, and individual controls) using data from 2011 to 2021 (all years affected by paywalls). While the relationship is somewhat noisy, respondents exposed to higher levels of paywalls clearly answer fewer knowledge questions correctly.

Table A3 confirms this pattern using regression analysis based on Equation (3), replacing treatment indicators with the continuous Paywall Exposure variable. Columns (1)–(3) include county and year fixed effects, while columns (4)–(6) add state-by-year fixed ef-

[15]This check also addresses the fact that *AAM* data are only available for counties with at least 25 sold copies.

Figure 8. Correlation of paywall exposure and political knowledge



Notes: Dots represent the average share of correct answers to eight political knowledge questions, residualized on individual characteristics as well as county- and year fixed effects, within each of 25 quantiles of paywall exposure. Sample weights are included. The black line shows the corresponding linear trend.

fects. Comparing Columns (1)–(2) and (4)–(5) reveals that the inclusion of individual demographic controls reduces the correlation between paywall exposure and knowledge while improving explanatory power. This result reflects both the correlation of paywall exposure with individual characteristics, and confirms that different types of individuals differ in their level of informedness about politics. However, comparing Columns (2)–(3) and (5)–(6) shows that interacting demographics with year fixed effects yields little additional explanatory power. This supports the central assumption that the negative correlation of paywalls and political knowledge is not systematically driven by unrelated shocks to demographic subgroups.

The results persist when relying only on within state-year variation, though slightly attenuated and only significant at the 10 percent level. This pattern may reflect that this specification is more restrictive, as paywall exposure varies more across than within states. Alternatively, higher paywall exposure may simply coincide with greater political change across states, making it more difficult to correctly guess majorities or party affiliations.

Table A4 replicates the previous analysis using bins corresponding to terciles of paywall exposure. Although estimates are mostly insignificant at conventional levels due to limited statistical power, the coefficient for the second tercile is consistently negative across specifications, and the coefficient for the third tercile is more negative and marginally significant in some cases. This pattern suggests that the results are not driven by outliers, and instead supports a monotonic negative relationship between paywall exposure and political knowledge.

Main result

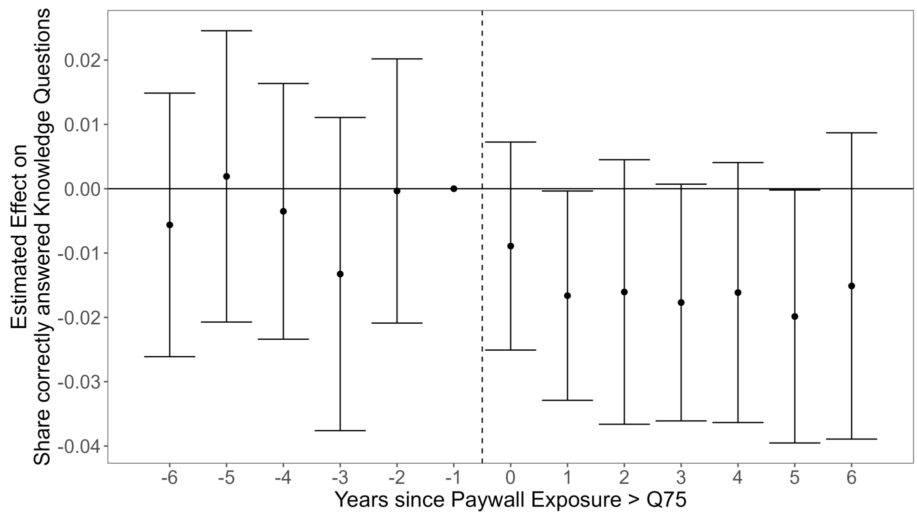
I now present the main results, motivated by the causal specification in Equation (3) and estimated via Callaway-Sant'Anna. Figure 9b displays the dynamic effect of paywalls on the share of correctly answered political knowledge questions, comparing survey respondents in counties with high versus low paywall exposure. Panel (a) includes county and year fixed effects, while Panel (b) adds state-by-year fixed effects. Both specifications confirm that the negative effect of paywall exposure emerges in the year of treatment and remains persistent after one to two years. Importantly, the estimates show no evidence of pre-treatment trends in political knowledge. Moreover, the similar-sized effect obtained using only within state-year variation helps rule out confounding from cross-state differences in the political environment.

Table 4 demonstrates that the result is robust to the choice of control variables and fixed effects. Column (1) reports results from a specification with county and year fixed effects and no individual controls, which corresponds to the most basic application of the Callaway-Sant'Anna estimator. Columns (2) and (3) first residualize the outcome on demographic controls, in levels and interacted with year fixed effects, respectively. These adjustments increase precision considerably, without affecting the magnitude of the estimate. Columns (4) and (5) repeat the estimation using state-by-year fixed effects, and yield similar conclusions. Together, the results suggest that, conditional on fixed effects and demographic controls, the bias from omitted variables is negligible. In particular, neither unrelated shocks to specific subpopulations nor potential measurement error in paywall exposure appear to drive the results.

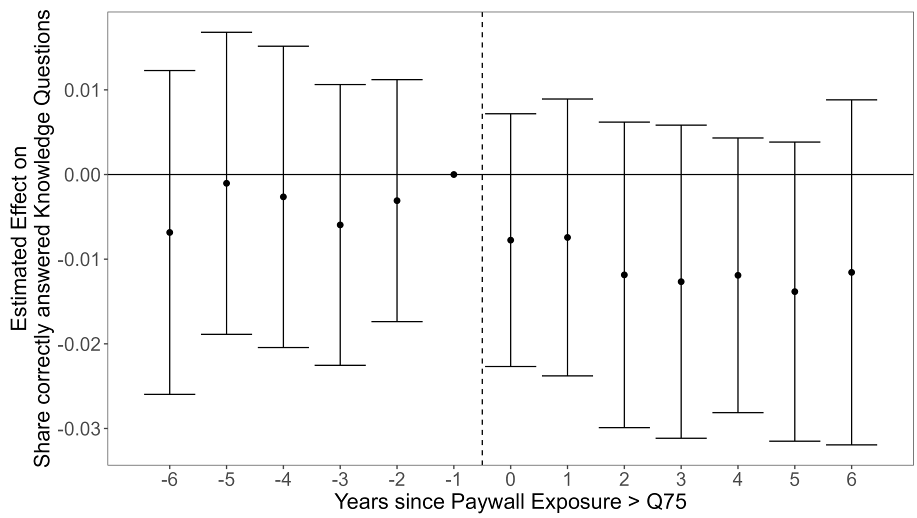
The most conservative estimate implies that high paywall exposure reduces the probability of correctly answering a political knowledge question by 1.2 percentage points (1.6 percent) relative to low exposure. For reference, this estimate corresponds to roughly 1 additional incorrect answer per 8 respondents, or 1 out of every 46 respondents now answering all questions incorrectly. Although modest in absolute terms, the effect is sizable given that only about half the sample regularly consumes news, and only a subset is expected to reduce consumption in response to paywalls. I return to the discussion of effect magnitudes at the end of this Section.

Figure 9. Effect of paywalls on political knowledge

(a) County and year fixed effects



(b) County and state-by-year fixed effects



Notes: Yearly estimates for the effect of high exposure to paywalls on political knowledge. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021), residualized on individual demographic controls, county fixed effects, and either year fixed effects (Panel (a)) or state-by-year fixed effects (Panel (b)). Counties are treated once paywall exposure exceeds the 75th percentile. The control group consists of not-yet-treated and never-treated counties. Vertical bars show bootstrapped 95 percent uniform confidence intervals. Sample weights are included. Standard errors are clustered by county.

Table 4. Effect of paywalls on political knowledge

Dependent Variable:	Share Correct Answers				
	(1)	(2)	(3)	(4)	(5)
Paywall Exposure > Q75	-0.018*	-0.017***	-0.018**	-0.012**	-0.015***
	(0.008)	(0.005)	(0.006)	(0.004)	(0.004)
Dep. Var. Mean	0.745	0.745	0.745	0.745	0.745
County FE	✓	✓	✓	✓	✓
Year FE	✓	✓		✓	
State-Year FE			✓		✓
Individual Controls		✓		✓	
Individual Controls x Year			✓		✓
Observations	530,372	518,867	518,867	518,867	518,867

Notes: Estimates for the effect of high exposure to paywalls on political knowledge. The specification is motivated by Equation (3), estimated and aggregated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021). All estimations include county and year fixed effects. Where indicated, the outcome is additionally residualized on individual demographic controls – categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2) – as well as county fixed effects and either year or state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Further robustness

The findings are robust to varying the threshold that defines whether counties are classified as as high- or low-paywall exposure. Figure A5 presents estimates for the baseline specification under different threshold levels. For easier comparison, the figure also reports the associated shares of survey respondents in the never-treated group. The effects remain quantitatively similar and at least marginally significant for a wide range of control group sizes.

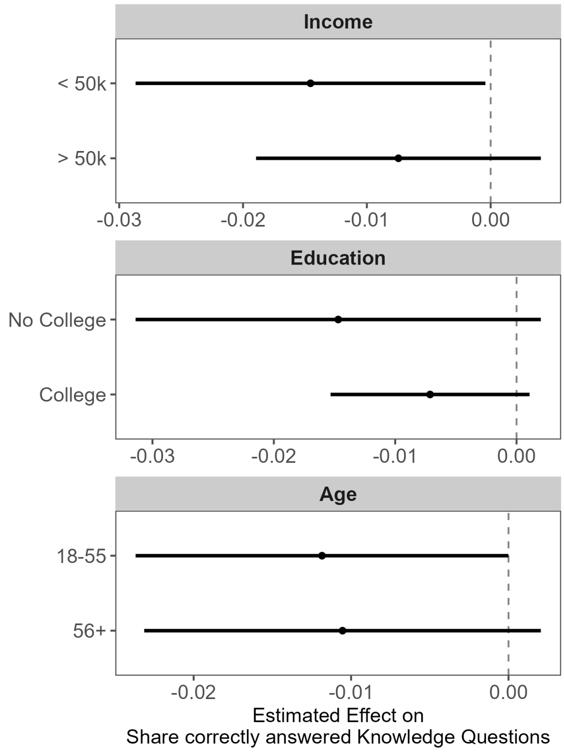
Heterogeneity

To better understand which types of political knowledge are affected by paywalls, I decompose the knowledge index by estimating eight separate regressions, each using as the outcome one of the index components. Figure A6 shows results, with Panel (a) reporting the specification with year fixed effects only, and Panel (b) including state-by-year fixed effects instead. Estimates in black indicate conventional 95 percent confidence intervals, while those in red apply Bonferroni adjustments that account (conservatively) for multiple hypothesis testing. The four upper rows correspond to questions about the majority party in state or federal legislative bodies, and the four lower rows pertain to questions regarding party affiliation given the name of a political representative.

The results show that the effects on political knowledge are driven by reduced knowledge of both party majorities and the party affiliations of elected representatives. This pattern is consistent across both specifications, though the estimates are somewhat attenuated in the more restrictive version. There is no clear evidence that knowledge of state-level politics declined more than knowledge of federal politics. Overall, these findings align most closely with a general decline in attention to politics.

Which types of individuals are most affected by paywalls? To answer this question, Figure 10 shows results from estimating the baseline specification separately for subgroups of survey respondents defined by family income, education, or age. Individuals with low income and low education levels experience larger declines in political knowledge. This finding is consistent with the notion that these individuals face relatively higher monetary costs of purchasing a paywall subscription, and may derive lower subjective benefits from high-quality information, either due to less interest in or a lower perceived value of such content. However, effects seem to be similar for older and younger individuals.

Figure 10. Effect of paywalls on political knowledge - Heterogeneity by individual characteristics



Notes: Estimates for the effect of high exposure to paywalls on political knowledge. Each panel reports coefficients from two separate estimations, corresponding to subgroups defined by yearly family income, reported education level, and age group, respectively. Horizontal bars show bootstrapped 95 percent confidence intervals. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021), residualized on individual demographic controls, county fixed effects, and state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75th percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors are clustered by county.

Mechanism

Up to this point, I have shown that paywalls reduce political knowledge. Two distinct mechanisms could explain this effect. First, individuals who previously accessed newspaper websites for free may be unwilling to pay for a subscription, and instead either reduce their consumption or switch to alternative sources with lower news content. While the previous analysis in Section 3 demonstrates that many readers do switch away from paywalled news sources, it is not immediately clear whether these individuals are also those experiencing declines in knowledge. Hence, a second possibility is that newspapers altered their content along with the introduction of paywalls – for instance, by reducing political coverage or overall quality. In that case, the effect could be driven by individuals who continued consuming the now-paywalled newspapers.

To address this limitation, I complement the analysis with survey data that records whether individuals pay for news subscriptions. Specifically, I use the 2018 *Pew–Ipsos Local News Survey* of more than 34,000 US adults. This survey includes a question on whether respondents paid for a subscription to local news source in the past year, providing information on the types of individuals that are more likely to pay for news.

I estimate individuals' propensities to pay for news using categorical indicators for age, sex, marital status, education, income, ethnicity, census region, and party identification. I implement four prediction models: linear regression, logistic regression, random forest, and gradient-boosted random forest (XGBoost). For each method, I select the best-performing estimator based on the F1 score, a common metric that balanced precision and recall in the positive (paying) group. I tune any hyperparameters via a successive halving random search across a manually defined grid of approximately 10,000 candidate configurations, evaluated using four-fold cross-validation. To account for sampling probabilities as well as the smaller share of subscription purchasers in the sample, I apply the survey's sampling weights multiplied by inverse class weights.

Figure A7 shows the benchmark distribution of F1 scores for each model. XGBoost, the best-performing model, achieves an average F1 score of approximately 0.50, highlighting the difficulty of the classification task. Nonetheless, the prediction yields highly informative results. Performance considerably and consistently exceeds an uninformative ("dummy") model, which scores only at around 0.23. Cross-validated precision is 43 percent for predicted subscribers and 89 percent for predicted non-subscribers, compared to baseline rates of 14 and 86 percent, respectively. This represents a substantial improvement, particularly for the smaller and more relevant group of paying users.

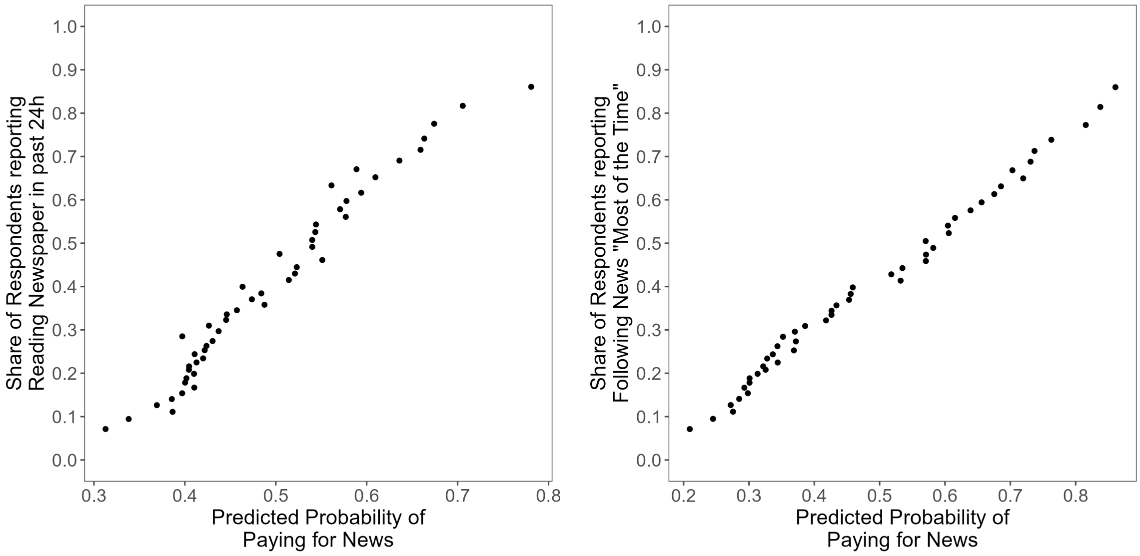
All predictors contribute to the model's performance, though with varying intensities. Figure A8 displays impurity-based feature importance scores, which measure the improvement in the score of the gradient-boosted random forest attributable to each variable. Age emerges as the most important predictor, explaining almost half of the model's predictive power. However, other variables variables also play an important role, most notably, education and income.

I use the XGBoost model trained on the *Pew-Ipsos* data to predict subscription probabilities in the *CES*. As an initial check of external validity, I examine their correlation with two measures that reveal intensive engagement with news. Figure 11 displays binned scatter plots relating predicted subscription probabilities to these outcomes. Panel (a) plots these probabilities against indicators for whether the respondent reports following the news "most of the time," the highest of four response categories. Panel (b) shows the relationship with indicators for whether the respondent read a newspaper in the past 24 hours. Both figures reveal strong, monotonic, and positive correlations across the full distribution of predicted probabilities. These patterns confirm that the predicted subscription probabilities capture meaningful variation in real-world news engagement, consistent with higher actual propensities to purchase subscriptions.

Figure 11. Correlation of predicted probability to pay for news with proxies for news interest

(a) Read newspaper in past 24h

(b) Follow news "most of the time"



Notes: Binned scatter plots of CES respondents (2006–2021). The x-axis shows the predicted probability of paying for a news subscription, based on the best XGBoost classifier trained on the Pew Research Center (2018) survey. The y-axis in Panel (a) is an indicator for whether the respondent reported reading a newspaper in the past 24 hours; in Panel (b), for whether the respondent reports following the news "most of the time," the highest among four categories. Sample weights are included.

Using the predicted subscription probabilities, I repeat the previous analysis for the effect of paywalls on political knowledge as motivated by Equation (3), this time splitting the sample by whether survey respondents are predicted to pay for news. Table A6 provides motivational evidence by correlating paywall exposure with political knowledge separately for the two groups. Columns (1)-–(2) report estimates for respondents predicted *not* to pay for news, while Columns (3)—(4) report results for those predicted to pay. The negative correlation between paywalls and knowledge is entirely driven by the non-paying group. Among predicted subscribers, by contrast, the correlation is statistically indistinguishable from zero, despite only a modest loss in precision.

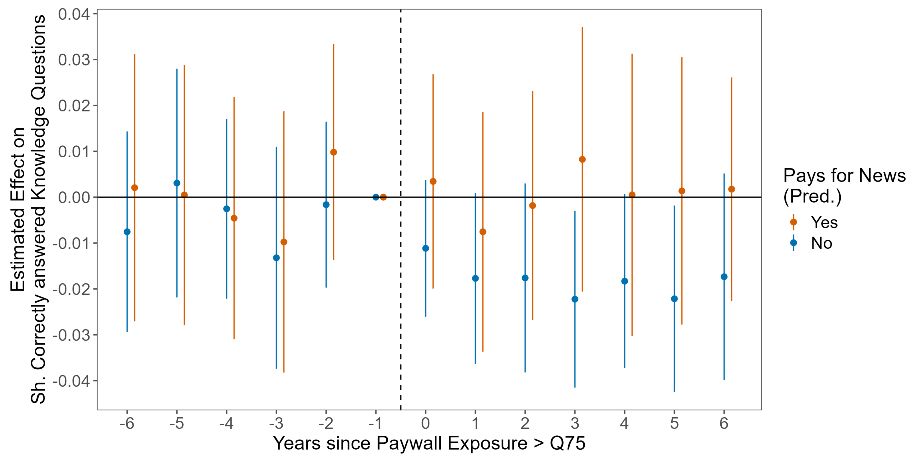
The event study estimates in Figure 12 confirm this pattern. The orange series corresponds to respondents predicted to pay for news, while the blue series reflects those predicted not to pay. The reduction in political knowledge is entirely driven by individuals predicted not to pay for news. Moreover, compared to the full sample, yearly effects are estimated with considerably higher precision. Instead, individuals predicted to pay for news are on a similar trajectory before treatment, but experience no changes in political knowledge after being exposed to paywalls.

Table 5 presents the corresponding average treatment effects. Among non-paying individuals, the estimated reduction in the share of correctly answered questions ranges from 1.5 to 1.9 percentage points, depending on the specification, which is slightly larger than the 1.2 to 1.8 percentage points estimated in the full sample. In contrast, for individuals predicted to pay for news, paywall exposure has no measurable effect on political knowledge in either specification.

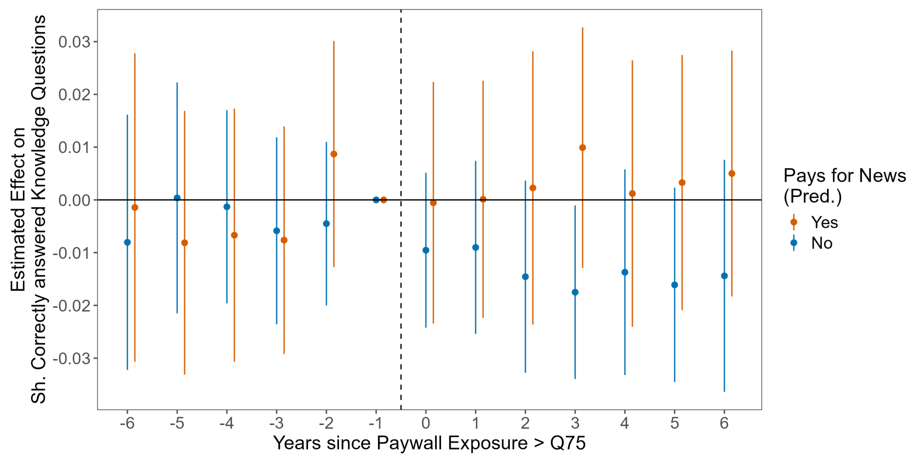
These findings reject the hypothesis that the observed declines in political knowledge are driven by changes in the news reporting of paywalled news websites, such as content or slant. If such changes were responsible, one would expect the effects to be concentrated among the individuals who continue to access and engage most intensively with the newspapers, which are most likely those who purchase subscriptions. Instead, the results are most consistent with the alternative hypothesis: Individuals unwilling to pay for news may reduce their consumption of paywalled newspapers and either switch to alternative sources with less political content or lower their news consumption overall. The resulting decline in attention to politics may then lead to the observed reduction in political knowledge.

Figure 12. Effect of paywalls on political knowledge - Heterogeneity by paying for news

(a) County and year fixed effects



(b) County and state-by-year fixed effects



Notes: Yearly estimates for the effect of high exposure to paywalls on political knowledge, separately by whether individuals are predicted to pay for a news subscription. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021), residualized on individual demographic controls, county fixed effects, and either year fixed effects (Panel (a)) or state-by-year fixed effects (Panel (b)). Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Vertical bars show bootstrapped 95 percent uniform confidence intervals. Sample weights are included. Standard errors are clustered by county.

Table 5. Effect of paywalls on political knowledge - Heterogeneity by paying for news

Dependent Variable:	Share Correct Answers			
Paying for News (Pred.):	No		Yes	
	(1)	(2)	(3)	(4)
Paywall Exposure > Q75	-0.019**	-0.015**	-0.001	0.002
	(0.006)	(0.005)	(0.007)	(0.007)
Dep. Var. Mean	0.635	0.635	0.842	0.842
County FE	✓	✓	✓	✓
Year FE	✓		✓	
State-Year FE		✓		✓
Individual Controls	✓	✓	✓	✓
Observations	413,797	413,797	105,070	105,070

Notes: Estimates for the effect of high exposure to paywalls on political knowledge, estimated separately by whether individuals are predicted to pay for a news subscription. The specification is motivated by Equation (3), estimated and aggregated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021). All estimations include county and year fixed effects. Where indicated, the outcome is additionally residualized on individual demographic controls – categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2) – as well as county fixed effects and either year or state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Magnitudes

In this subsection I discuss the magnitude of the estimated effect of paywalls on political knowledge. To facilitate comparisons with other studies in the media literature, I present a back-of-envelope calculation for the "persuasion rate" of paywalls, defined as the share of individuals who switched from the correct to the incorrect answer because of paywalls, among those exposed to paywalls.[16] Following the notation of DellaVigna and Gentzkow (2010), the persuasion rate for this setting is calculated as:

$$f = \frac{y_T - y_C}{e_T - e_C} \frac{1}{1 - y_0} \tag{4}$$

$$= \frac{0.012}{0.5 * 0.5} \frac{1}{0.745} = 0.048 \tag{5}$$

Here, $y_T - y_C$ is the change in the outcome between the treated and the control group in response to the treatment. To provide a conservative estimate, I rely on the smallest coefficient for the main effect which is 0.012. $e_T - e_C$ denotes the difference in the share of the population exposed to the treatment. I assume that around 25 percent of the voting-age population is exposed, resulting from the observations that around 50 percent of the survey respondents read newspapers online during the sample period, and that average

[16] In this setting, the term "persuasion rate" applies imperfectly, as individuals do not change their *opinion* in response to the treatment.

paywall exposure is around twice as high in the treated group compared to the control group. $1 - y_0$ represents the share of the population left to be persuaded. In this context, I use the average share of correct questions in the population at baseline, which is 0.745 (calculated using the years 2006 through 2010).

The resulting magnitude is 4.8 percent. On average, this means that approximately 4.8 percent among individuals who initially read a newspaper that implemented a paywall lost all their political knowledge of the type tested in the survey. As an equivalent interpretation, 9.6 percent answered only 3 instead of 6 out of the eight questions correctly, or only half as many as initially.

The estimated persuasion rate of 4.8 is in the lower range of estimates in the literature on the effects of news availability on political outcomes (DellaVigna and Gentzkow, 2010). For example, Enikolopov et al. (2011) find a persuasion rate of 7.7 percent from the availability of the independent anti-Putin TV station *NTV* on reducing the vote share for pro-Putin parties. Similarly, DellaVigna and Kaplan (2007) estimate an 11.6 percent persuasion rate from the availability of *Fox News* on the Republican vote share in presidential elections. Gerber et al. (2009) find a rate of 19.5 percent for free subscriptions to the *Washington Post* increasing intent to vote Democratic. Gentzkow et al. (2011) find a 12.9 percent persuasion rate for local newspaper readership in the late 1800s and early 1900s on voter turnout in presidential elections, while Gentzkow (2006) reports a 4.4 percent persuasion rate for early television exposure reducing congressional election turnout.

While informative, my calculation has some limitations. First, since I measure knowledge as an index comprised of several outcomes, the estimates reflect the effect of paywalls on the share of answers in the population, rather than on the number of individuals who lose knowledge. Therefore, the calculation implicitly assumes that the effect of paywalls on knowledge is homogeneous across individuals. Consequently, if reductions knowledge are concentrated among fewer individuals, the true persuasion rate is higher; if many individuals lose a smaller amount of knowledge, the rate is lower.

Second, the calculation should be interpreted as the intent-to-treat effect of being exposed to a paywall. It represents the net effect of two opposing mechanisms: increases in consumption by some individuals who purchase a subscription, and decreases in consumption by others who substitute away. Without more detailed data to disentangle these mechanisms, it is not possible to compute a persuasion rate specifically for the subset of readers who reduce their consumption due to paywalls.

4.5 Effects on electoral participation

Setup

To estimate the effects of paywalls on electoral participation, I use voting data from the *CES*, which is available in two forms: self-reported and validated. My primary outcome is participation in major elections – specifically, for president, governor, senators, and federal House representative. These measures are available in both self-reported and validated formats, enabling direct comparison. I construct an index of vote participation by calculating, for each respondent, the share of elections in which they participated, based on the elections held in their state and year. That way, the approach captures both the extensive and intensive margins of participation.

Because elections occur only every two years or less, I adapt the specification from Equation (3) to a two-year frequency, defining treatment status based on paywall exposure as of October in each election year. As in prior analyses, I apply the Callaway–Sant'Anna estimator, residualizing on individual characteristics as well as county and state-by-year fixed effects. Note that the latter are preferred because they effectively control for the number of elections in each state and year, which directly affects the outcome.

Main results

Table 6 reports results for different population subsets. Panel A uses self-reported turnout, and Panel B uses validated turnout. Columns (1)–(2) present estimates for the full sample. Including state-by-year fixed effects in Column (2) is important, as it absorbs variation in election exposure across states and years. While the estimated effects are negative, they are not statistically significant.

Motivated by the earlier finding that paywall effects on political knowledge are concentrated among individuals predicted not to pay for news, Column (3) restricts the sample to this group. Consistent with prior results, the estimated effects are larger in magnitude for both self-reported and validated participation. Moreover, they are statistically significant for self-reported participation.

Finally, Column (4) further restricts the sample by excluding the most partisan counties, defined as those with a partisanship index between 0.3 and 0.7 (where 0.5 indicates a perfectly balanced electorate between 2002 and 2010). These more competitive counties are particularly relevant, as they feature stronger incentives for individuals to stay informed and are more likely to exhibit meaningful electoral consequences from shifts in participation. In this subsample, the estimated effects become more pronounced for both self-reported and validated outcomes. The effect on self-reported participation is highly significant, while the effect on validated participation is now significant at the 10 percent level.

A comparison between self-reported and validated voting data reveals evidence for misreporting. On average, self-reported turnout exceeds validated participation by approximately 10 percentage points. As noted by Ansolabehere and Hersh (2012), this discrepancy most likely reflects intentional overreporting driven by experimenter demand effects, although other factors such as faulty recall or errors in the validation procedure may also play a role. While such potential measurement error in either variable complicates the interpretation of absolute values, it affects my analysis only if *changes* in mismeasurement are correlated with paywall exposure, conditional on fixed effects and controls. In this respect, it is reassuring that effect sizes relative to the respective baseline are similar across both measures. Among respondents predicted not to pay for news, the average reduction in participation is roughly 2 percent for both self-reported and validated turnout.

Figure 13 presents event study estimates for the sample used in Column (4), restricted to non-paying individuals in less partisan counties. Panel (a) displays results for self-reported voting, while Panel (b) shows the corresponding estimates for validated participation. Both variables exhibit no discernible pre-treatment trends, but reveal a decline in participation during the first two post-treatment periods that remains stable afterwards. This pattern closely mirrors the event study results for political knowledge, supporting the interpretation that reduced information and lower turnout are connected. Taken together, the findings suggest that paywalls, by lowering political engagement, ultimately reduce participation in the democratic process.

Heterogeneity

Figure A10 and Table A7 present a decomposition of the vote participation index by election type. I estimate separate regressions using as the outcome a binary indicator for participation in each election. I restrict the sample to individuals predicted not to pay for news, and exclude the most partisan counties. The results suggest that the decline in participation is broadly consistent across all four election types. Interestingly, the self-reported decline is primarily driven by elections for governor and house representative, whereas for the validated data, the effects are largest for congressional races.

In addition to these core outcomes, I examine participation for other elections available in the survey. For the indicator of whether the respondent participated in any general election, point estimates are negative but only marginally insignificant for both self-reported

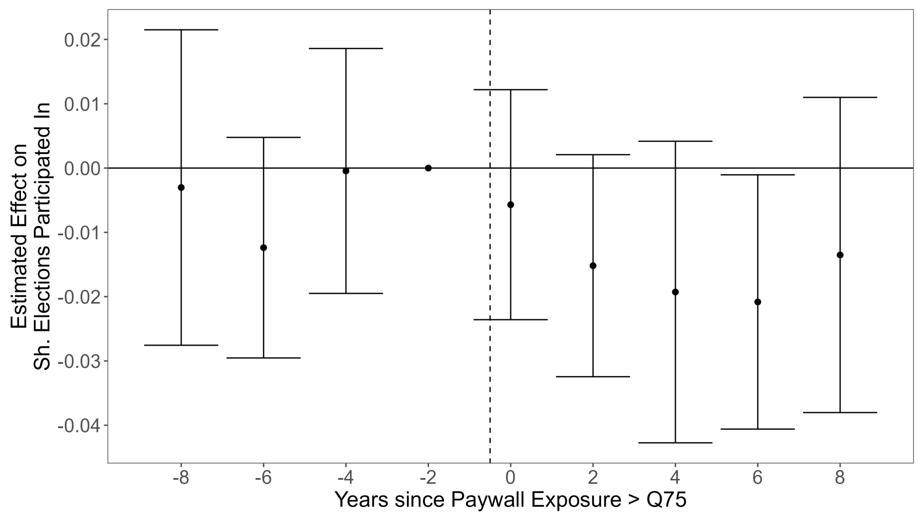
Table 6. Effect of paywalls on electoral participation

Dependent Variable:	Participation in Elections			
Paying for News (Pred.):	All		No	
County Partisanship Index:	All			0.3-0.7
	(1)	(2)	(3)	(4)
Panel A: Self-Reported				
Paywall Exposure > Q75	-0.0092	-0.0056	-0.0119**	-0.0147***
	(0.0068)	(0.0049)	(0.0060)	(0.0055)
Dep. Var. Mean	0.6072	0.6072	0.5736	0.5739
County FE	✓	✓	✓	✓
Year FE	✓			
State-Year FE		✓	✓	✓
Individual Controls	✓	✓	✓	✓
Observations	418,328	418,328	330,193	279,457
Panel B: Validated				
Paywall Exposure > Q75	-0.0142**	-0.0055	-0.0090	-0.0115*
	(0.0064)	(0.0051)	(0.0057)	(0.0063)
Dep. Var. Mean	0.5089	0.5089	0.4798	0.4814
County FE	✓	✓	✓	✓
Year FE	✓			
State-Year FE		✓	✓	✓
Individual Controls	✓	✓	✓	✓
Observations	383,040	383,040	300,286	253,940

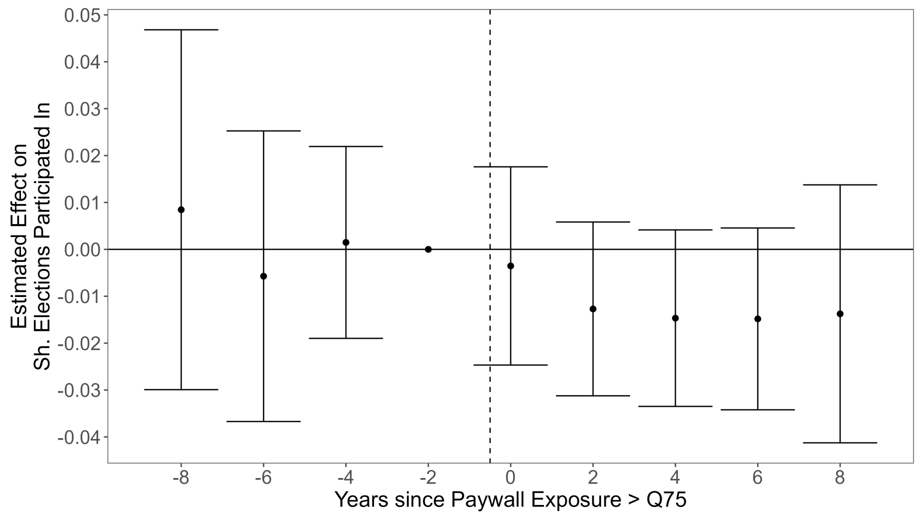
Notes: Estimates for the effect of high exposure to paywalls on electoral participation. The specification is motivated by Equation (3), estimated and aggregated using the Callaway-Sant'Anna estimator. The outcome is the share of elections participated in among elections held for president, house representative, senators, and governor (CES, 2008–2020). Panel A uses self-reported voting; Panel B uses validated data from matching survey respondents to official records (Ansolabehere and Hersh, 2012). All specifications residualize the outcome on individual demographic controls – categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2) – as well as county fixed effects. Column (1) adds year fixed effects, and Columns (2)–(4) add state-by-year fixed effects. Columns (1)–(2) use the full sample. Column (3) restricts to individuals predicted not to pay for news. Column (4) further restricts to counties with a 2010 partisanship index between 0.3 and 0.7, (see Chenoweth et al., 2020). Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 13. Effect of paywalls on participation in elections - Individuals not paying for news, Non-partisan counties

(a) Self-reported



(b) Validated



Notes: Bi-yearly estimates for the effect of high exposure to paywalls on electoral participation, among individuals predicted not to pay for a news subscription and residing counties with low partisanship. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of elections participated in among elections held for president, house representative, senators, and governor (CES, 2008–2020), residualized on individual demographic controls, county fixed effects, and state-by-year fixed effects. Panel (a) uses self-reported voting data, while Panel (b) uses validated data from matching survey respondents to official voting records (Ansolabehere and Hersh, 2012). Counties are treated once paywall exposure exceeds the 75th percentile. The control group consists of not-yet-treated and never-treated counties. Vertical bars show bootstrapped 90 percent uniform confidence intervals. Sample weights are included. Standard errors are clustered by county.

and validated data. Notably, self-reported participation in state House and Senate elections also declines and exhibits larger magnitudes than other outcomes. This pattern may suggest somewhat stronger effects on state-level participation, potentially reflecting the fact that local newspapers, which are more likely to cover state politics, were disproportionately affected by paywalls relative to national outlets. Overall, these findings draw a consistent picture of negative effects of paywalls on electoral participation.

5 Conclusion

In this paper, I demonstrate that the introduction of paywalls on US newspaper websites reduced political knowledge and electoral participation. Paywalls led to substantial declines in online consumption of traditional newspapers. As a result, regions highly exposed to paywalls experienced declines in factual knowledge about contemporary politics, especially among less-educated and low-income individuals. These findings suggest that such groups are less willing or able to pay for digital news, leaving them disproportionately uninformed. In turn, paywalls lowered electoral participation among these populations, especially in politically competitive regions.

My findings underscore that news media play a critical role for the democratic process. In a media landscape dominated by private, profit-oriented news outlets, economic shocks can require business decision that ultimately affect voters' information and participation. Therefore, to counteract the potentially adverse political effects of such economic shifts, modern democracies should ensure that broad and equitable access to relevant information.

To better understand both the mechanisms behind my results as well as potential remedies, further research could explore how readers substitute paywalled websites, and examine supply-side changes to topics, slant, or quality that may accompany paywalls. Moreover, one could further investigate downstream effects on political opinions and party preference, particularly in local elections. Finally, it would be valuable to determine whether the displacement of readers from traditional media contributed to the rise of right-wing populism in the following years.

References

Angelucci, Charles and Julia Cagé, "Newspapers in times of low advertising revenues," *American Economic Journal: Microeconomics*, 2019, *11* (3), 319–64.

Ansolabehere, Stephen and Brian Schaffner, "CES Common Content, 2021," 2022.

— **and Eitan Hersh**, "Validation: What big data reveal about survey misreporting and the real electorate," *Political Analysis*, 2012, *20* (4), 437–459.

Aral, Sinan and Paramveer S Dhillon, "Digital paywall design: Implications for content demand and subscriptions," *Management Science*, 2021, *67* (4), 2381–2402.

Ash, Elliott and Sergio Galletta, "How cable news reshaped local government," *American Economic Journal: Applied Economics*, 2023, *15* (4), 292–320.

Baker, Andrew C, David F Larcker, and Charles CY Wang, "How much should we trust staggered difference-in-differences estimates?," *Journal of Financial Economics*, 2022, *144* (2), 370–395.

Bartik, Timothy J, "Who benefits from state and local economic development policies?," 1991.

Besley, Timothy and Robin Burgess, "The political economy of government responsiveness: Theory and evidence from India," *The quarterly journal of economics*, 2002, *117* (4), 1415–1451.

Bhuller, Manudeep, Tarjei Havnes, Jeremy McCauley, and Magne Mogstad, "How the internet changed the market for print media," *American Economic Journal: Applied Economics*, 2024, *16* (2), 318–358.

Blanchard, Olivier Jean, Lawrence F Katz, Robert E Hall, and Barry Eichengreen, "Regional evolutions," *Brookings papers on economic activity*, 1992, *1992* (1), 1–75.

Cagé, Julia, "Media competition, information provision and political participation: Evidence from French local newspapers and elections, 1944–2014," *Journal of Public Economics*, 2020, *185*, 104077.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of econometrics*, 2021, *225* (2), 200–230.

Campante, Filipe, Ruben Durante, and Francesco Sobbrio, "Politics 2.0: The multifaceted effect of broadband internet on political participation," *Journal of the European Economic Association*, 2018, *16* (4), 1094–1136.

Chenoweth, Megan, Mao Li, Iris N Gomez-Lopez, and Ken Kollman, "National Neighborhood Data Archive (NaNDA): Voter Registration, Turnout, and Partisanship by County, United States, 2004-2018," *Ann Arbor, MI: Inter-university Consortium for Political and Social Research[distributor]*, 2020, pp. 11–04.

Chiou, Lesley and Catherine Tucker, "Paywalls and the demand for news," *Information Economics and Policy*, 2013, *25* (2), 61–69.

Chung, Doug J, Ho Kim, and Reo Song, *The comprehensive effects of a digital paywall sales strategy*, Harvard Business School, 2019.

Cook, Jonathan E and Shahzeen Z Attari, "Paying for what was free: Lessons from the New York Times paywall," *Cyberpsychology, behavior, and social networking*, 2012, *15* (12), 682–687.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

— **and Matthew Gentzkow**, "Persuasion: empirical evidence," *Annu. Rev. Econ.*, 2010, *2* (1), 643–669.

Djourelova, Milena, Ruben Durante, and Gregory J Martin, "The impact of online competition on local newspapers: Evidence from the introduction of Craigslist," *Review of Economic Studies*, 2024, p. rdae049.

Drago, Francesco, Tommaso Nannicini, and Francesco Sobbrio, "Meet the press: How voters and politicians respond to newspaper entry and exit," *American Economic Journal: Applied Economics*, 2014, *6* (3), 159–88.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei, "The political legacy of entertainment TV," *American Economic Review*, 2019, *109* (7), 2497–2530.

Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya, "Media and political persuasion: Evidence from Russia," *American Economic Review*, 2011, *101* (7), 3253–85.

Falck, Oliver, Robert Gold, and Stephan Heblich, "E-lections: Voting Behavior and the Internet," *American Economic Review*, 2014, *104* (7), 2238–65.

Ferraz, Claudio and Frederico Finan, "Exposing corrupt politicians: the effects of Brazil's publicly released audits on electoral outcomes," *The Quarterly journal of economics*, 2008, *123* (2), 703–745.

Gao, Pengjie, Chang Lee, and Dermot Murphy, "Financing dies in darkness? The impact of newspaper closures on public finance," *Journal of Financial Economics*, 2020, *135* (2), 445–467.

Gavazza, Alessandro, Mattia Nardotto, and Tommaso Valletti, "Internet and politics: Evidence from UK local elections and local government policies," *The Review of Economic Studies*, 2019, *86* (5), 2092–2135.

Gentzkow, Matthew, "Television and voter turnout," *The Quarterly Journal of Economics*, 2006, *121* (3), 931–972.

_ and Jesse M Shapiro, "Introduction of Television to the United States Media Market, 1946-1960," 2008.

_ , _ , and Michael Sinkinson, "The effect of newspaper entry and exit on electoral politics," *American Economic Review*, 2011, *101* (7), 2980–3018.

_ , _ , and _ , "Competition and ideological diversity: Historical evidence from us newspapers," *American Economic Review*, 2014, *104* (10), 3073–3114.

George, Lisa M and Joel Waldfogel, "The New York Times and the market for local newspapers," *American Economic Review*, 2006, *96* (1), 435–447.

Gerber, Alan S, Dean Karlan, and Daniel Bergan, "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics*, 2009, *1* (2), 35–52.

Guriev, Sergei, Nikita Melnikov, and Ekaterina Zhuravskaya, "3g internet and confidence in government," *The Quarterly Journal of Economics*, 2021, *136* (4), 2533–2613.

Jr, James M Snyder and David Strömberg, "Press coverage and political accountability," *Journal of political Economy*, 2010, *118* (2), 355–408.

Kim, Ho, Reo Song, and Youngsoo Kim, "Newspapers' Content Policy and the Effect of Paywalls on Pageviews," *Journal of Interactive Marketing*, 2020, *49*, 54–69.

MRI-Simmons, "NCS Adult Study 12 Month," 2014.

Pattabhiramaiah, Adithya, S Sriram, and Puneet Manchanda, "Paywalls: Monetizing online content," *Journal of marketing*, 2019, *83* (2), 19–36.

Pew Research Center, "The State of the News Media 2013: An Annual Report on American Journalism," 2013. Accessed on January 23, 2025.

__ , “For Local News, Americans Embrace Digital but Still Want Strong Community Connection,” 2018. Accessed on April 28, 2025.

__ , “State of the News Media 2023,” 2023. Accessed on January 23, 2025.

Sant’Anna, Pedro HC and Jun Zhao, “Doubly robust difference-in-differences estimators,” *Journal of econometrics*, 2020, *219* (1), 101–122.

Seamans, Robert and Feng Zhu, “Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers,” *Management Science*, 2014, *60* (2), 476–493.

Strömberg, David, “Radio’s impact on public spending,” *The Quarterly Journal of Economics*, 2004, *119* (1), 189–221.

Sun, Liyang and Sarah Abraham, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 2021, *225* (2), 175–199.

U.S. Census Bureau, “American Community Survey 5-Year Estimates: Comparison Profiles 5-Year,” 2010. Accessed on April 22, 2024.

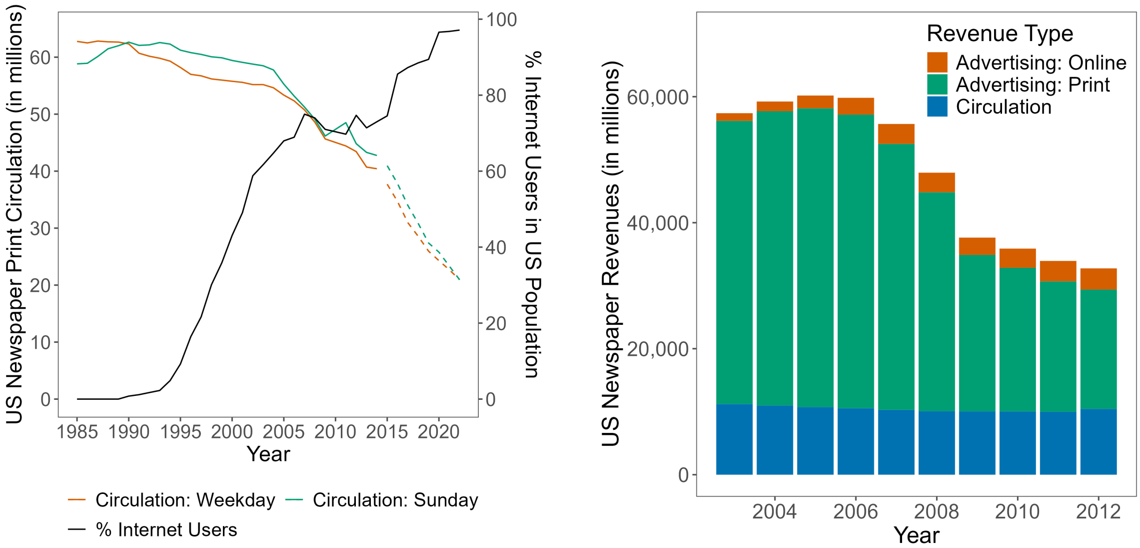
World Bank, “Internet users for the United States [ITNETUSERP2USA],” 2025. Retrieved from FRED, Federal Reserve Bank of St. Louis.

A Appendix

Appendix Figure A1. State of the US newspaper market in the 2000s

(a) Print circulation and share of internet users

(b) Revenues by type



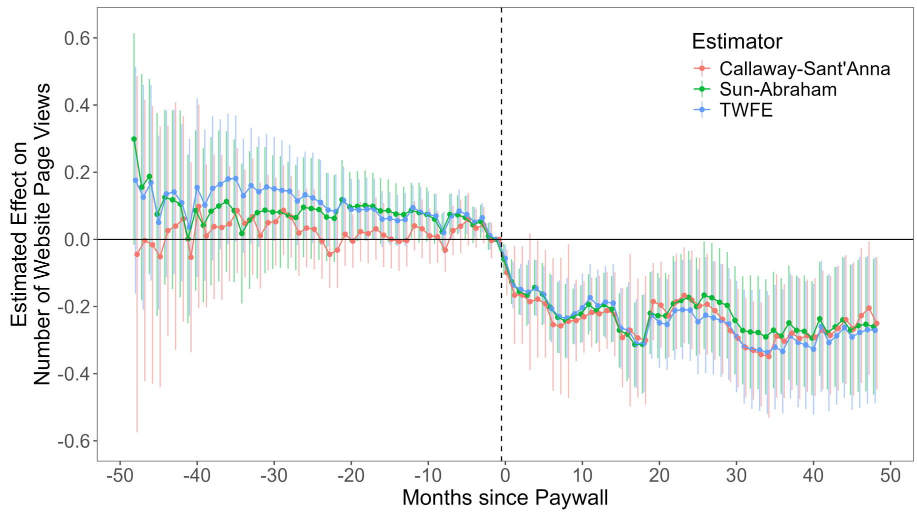
Notes: Panel (a) shows the total number of print copies sold of US daily newspapers by year. The orange and green line represent the circulation of the newspapers' weekday and Sunday editions, respectively. Dotted lines are estimates (Pew Research Center, 2023). The black line represents the share of the US population with access to the internet (World Bank, 2025). Panel (b) shows the composition of revenues among US daily newspapers by year. Orange and green bars represent revenues from online and print advertising, respectively, while the blue bars represent revenues from the sale of physical and virtual newspaper copies (Pew Research Center, 2013).

Appendix Table A1. Effect of paywalls on pageviews - Estimators

Dependent Variable:	Log(Pageviews)					
Estimator:	TWFE		Sun-Abraham		Callaway-Sant'Anna	
	(1)	(2)	(3)	(4)	(5)	(6)
Paywall	-0.258***	-0.273***	-0.267***	-0.279***	-0.257***	-0.280***
	(0.059)	(0.060)	(0.074)	(0.074)	(0.069)	(0.065)
Newspaper FE	✓	✓	✓	✓	✓	✓
Month-Year FE	✓	✓	✓	✓	✓	✓
Market Controls		✓		✓		✓
Observations	6,912	6,912	6,912	6,912	6,912	6,912
R^2	0.957	0.960	0.978	0.981		

Notes: Coefficients for static (pre-post) version of Equation (1), estimated via Ordinary Least Squares (TWFE), Sun-Abraham, or Callaway-Sant'Anna: Regressions of monthly number of newspaper website views on indicator denoting active paywall, controlling for newspaper fixed effects and year-month fixed effects. Columns (2), (4), and (6) additionally control for average county characteristics at 2010-levels where newspaper is active, weighted by number of print copies sold, interacted with year-month fixed effects: Log population density, share of households with yearly income below \$50,000 and above \$100,000, share adults with college degree, and NaNDA partisanship index. Standard errors in parentheses are clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figure A2. Effect of paywalls on website page views - Estimators



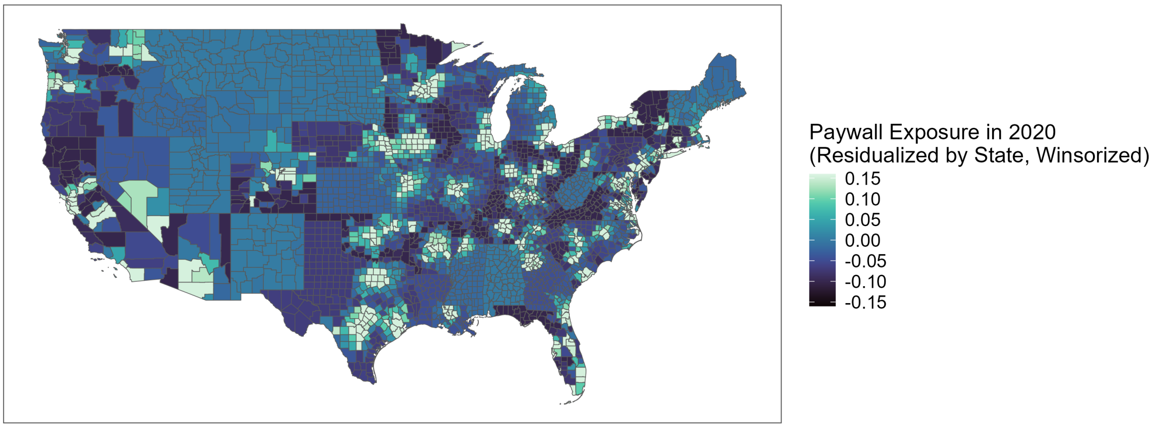
Notes: Monthly coefficients for β_τ from Equation (1), estimated either via Callaway-Sant'Anna, Sun-Abraham, or OLS (TWFE): Regression of monthly number of newspaper website views on indicators denoting the number of months since paywall implementation. The omitted category is the month before the paywall. Includes newspaper fixed effects, year-month fixed effects, and log population density, shares of three income buckets, college-educated share, and partisanship index for newspapers' audiences, interacted with year-month fixed effects. Vertical bars denote 95 percent pointwise confidence intervals. Standard errors are clustered by newspaper.

Appendix Table A2. Effect of paywalls on pageviews - Restrict sample

Dependent Var.:	Log(Pageviews)				
	Exclude Newspapers				Add Weights
	National	Top 15	Top 15 Paywalled	Bottom 5 Control	Weights
	(1)	(2)	(3)	(4)	(5)
Paywall	-0.282***	-0.332***	-0.357***	-0.320***	-0.273***
	(0.065)	(0.071)	(0.073)	(0.075)	(0.068)
Newspaper FE	✓	✓	✓	✓	✓
Month-Year FE	✓	✓	✓	✓	✓
Market Controls	✓	✓	✓	✓	✓
Restricted Control Group			✓	✓	
Sample Weights					Log(Pageviews 2010)
Observations	6,624	5,472	5,472	6,432	6,912

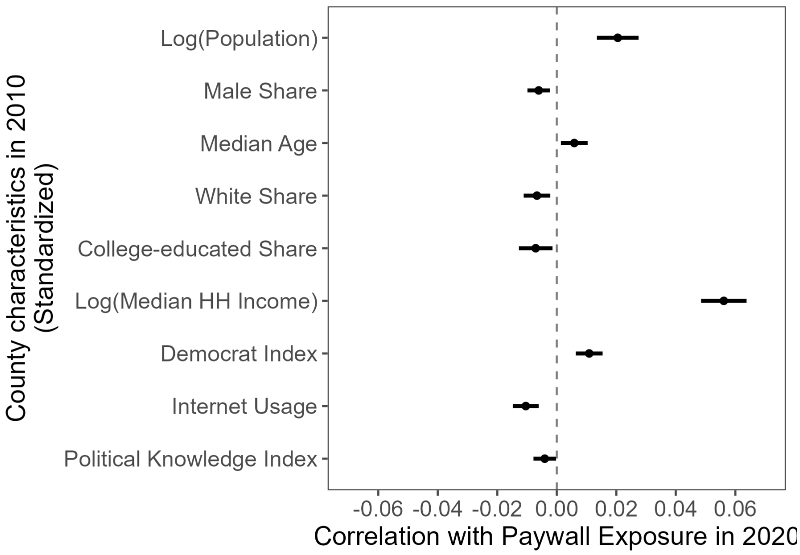
Notes: Coefficients for static (pre-post) version of Equation (1), excluding different subsets of newspapers from the sample. Column (1) excludes the three national newspapers *New York Times*, *USA Today*, and *Wall Street Journal*. Column (2) excludes the largest 15 newspapers in 2010 by website page views. Columns (3) and (4) exclude the largest 15 newspapers in the treated group, and the bottom 5 newspapers in the control group, respectively. In both columns, treated observations pre-treatment are also excluded from the control group. Column (5) includes the log of 2010 page views as sample weights. Regressions of monthly number of newspaper website views on indicator denoting active paywall. All specifications include newspaper fixed effects and year-month fixed effects. Controls represent average county characteristics at 2010-levels where newspaper is active, weighted by number of print copies sold, interacted with year-month fixed effects: Log population density, share of households with yearly income below \$50,000 and above \$100,000, share adults with college degree, and NaNDA partisanship index. Estimated using Callaway-Sant'Anna estimator. Standard errors in parentheses are clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figure A3. Geographical variation of paywall exposure - Residualized



Notes: Paywall exposure by county in 2020, residualized by state and winsorized at 10 and 90 percent.

Appendix Figure A4. County-level predictors of paywall exposure



Notes: Multivariate regression of county-level paywall exposure in 2020 on county characteristics in 2010. All variables are standardized. Horizontal bars denote 95 percent confidence intervals.

Appendix Table A3. Correlation of (continuous) paywall exposure and political knowledge

Dependent Variable:	Share Correct Answers					
	(1)	(2)	(3)	(4)	(5)	(6)
Paywall Exposure	-0.047***	-0.025**	-0.022**	-0.024*	-0.016*	-0.016*
	(0.013)	(0.011)	(0.011)	(0.013)	(0.009)	(0.010)
Dep. Var. Mean	0.745	0.745	0.745	0.745	0.745	0.745
County FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓			
State-Year FE				✓	✓	✓
Indiv. Demographics		✓			✓	
Indiv. Demographics x Year FE			✓			✓
Observations	529,409	517,924	517,924	529,409	517,924	517,924
R^2	0.080	0.329	0.320	0.091	0.338	0.328

Notes: Estimates from regressing the share of correct answers to eight political knowledge questions on paywall exposure. All specifications control for fixed effects as indicated. Individual demographic controls include categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2). Data are from the CES 2006–2021. Survey weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

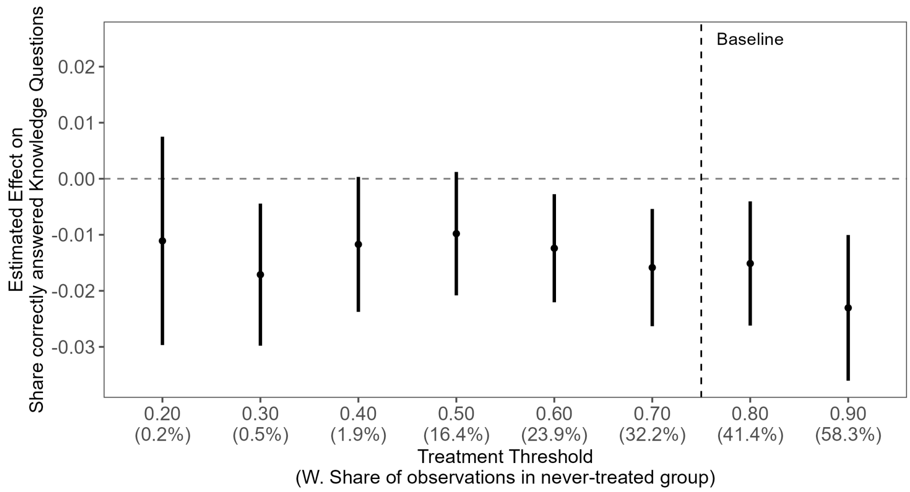
Appendix Table A4. Correlation of (binned) paywall exposure and political knowledge

Dependent Variable:	Share Correct Answers					
	(1)	(2)	(3)	(4)	(5)	(6)
Paywall Exposure, 2nd Tercile	-0.001	-0.004	-0.002	-0.002	-0.003	-0.004
	(0.005)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)
Paywall Exposure, 3rd Tercile	-0.009	-0.008*	-0.007	-0.007	-0.008	-0.009*
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Dep. Var. Mean	0.745	0.745	0.745	0.745	0.745	0.745
County FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓			
State-Year FE				✓	✓	✓
Indiv. Demographics		✓			✓	
Indiv. Demographics x Year FE			✓			✓
Observations	437,835	437,039	437,039	437,835	437,039	437,039
R^2	0.066	0.326	0.312	0.075	0.334	0.320

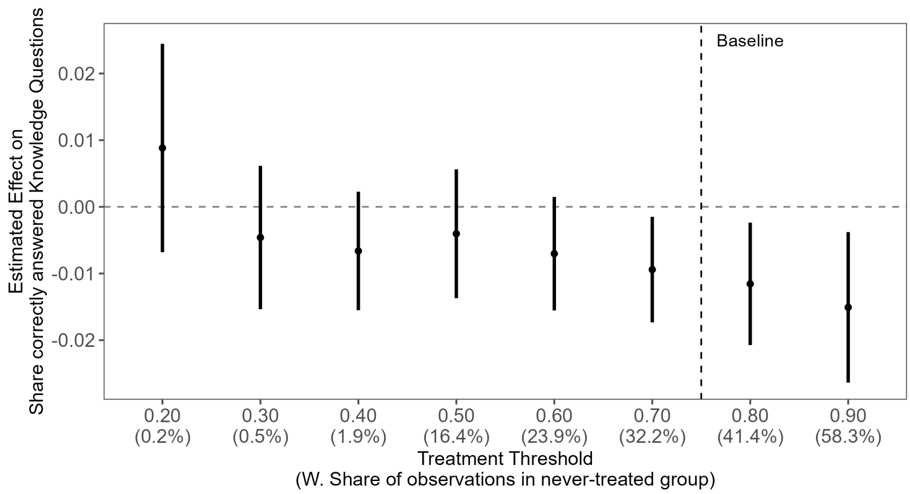
Notes: Estimates from regressing the share of correct answers to eight political knowledge questions on indicators for terciles of paywall exposure. All specifications control for fixed effects as indicated. Individual demographic controls include categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2). Data are from the CES 2006–2021. Survey weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figure A5. Effect of paywalls on political knowledge - Thresholds

(a) County and year fixed effects



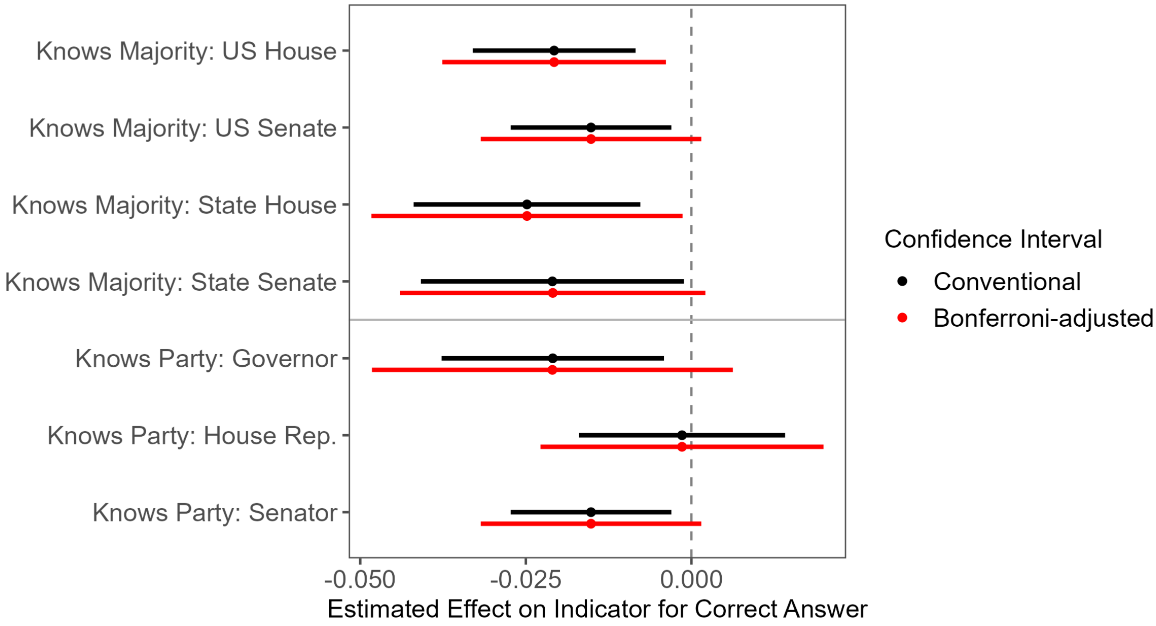
(b) County and state-by-year fixed effects



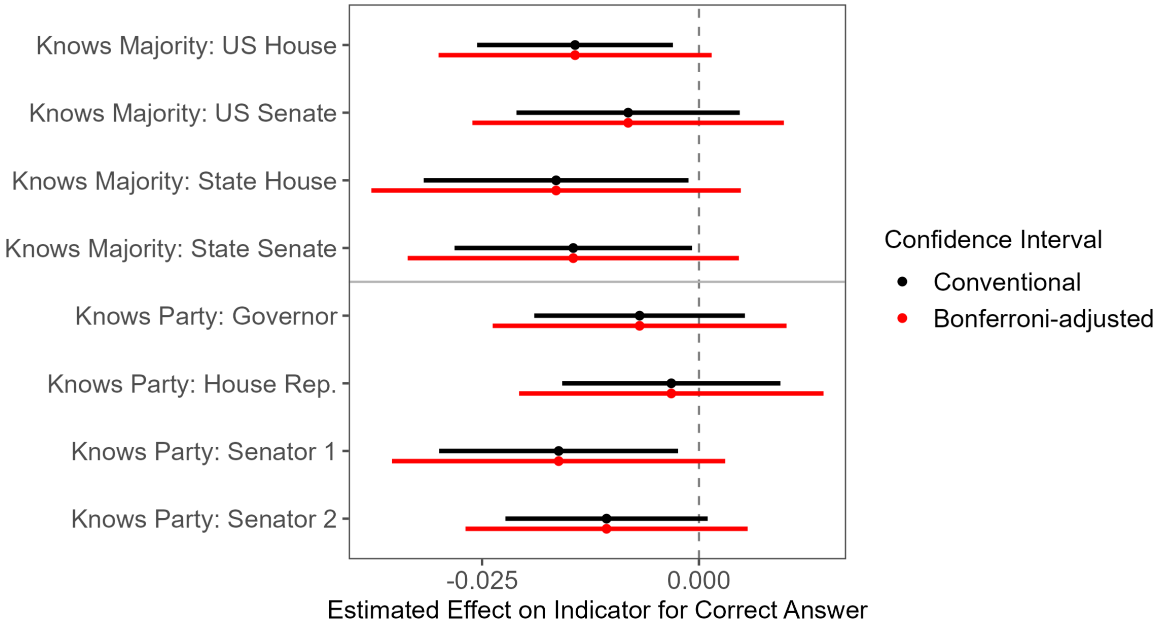
Notes: Estimates for the effect of high exposure to paywalls on political knowledge, for different percentile thresholds used to define when a county switches from low to high paywall exposure. The numbers in brackets indicate the sample-weighted share of observations in the never-treated group, across all years. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021), residualized on individual demographic controls, county fixed effects, and either year fixed effects (Panel (a)) or state-by-year fixed effects (Panel (b)). Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Vertical bars show bootstrapped 95 percent confidence intervals. Sample weights are included. Standard errors are clustered by county.

Appendix Figure A6. Effect of paywalls on political knowledge - Index decomposition

(a) County and year fixed effects



(b) County and state-by-year fixed effects



Notes: Each row represents an estimate for the effect of high exposure to paywalls on correctly answering the respective political knowledge question. Horizontal bars show bootstrapped 95 percent confidence intervals, estimated conventionally (black) or with Bonferroni adjustment (red). The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is an indicator for a correct answer (CES, 2006–2021), residualized on individual demographic controls, county fixed effects, and either year fixed effects (Panel (a)) or state-by-year fixed effects (Panel (b)). Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors are clustered by county.

Appendix Table A5. Effect of paywalls on political knowledge - Heterogeneity by individual characteristics

Dependent Variable:	Share Correct Answers					
Subsample:	Income		Education		Age	
	$\leq$ 50k	> 50k	No College	College	18-55	56+
	(1)	(2)	(3)	(4)	(5)	(6)
Paywall Exposure > Q75	-0.015**	-0.007	-0.015*	-0.007*	-0.012**	-0.011
	(0.007)	(0.006)	(0.009)	(0.004)	(0.006)	(0.006)
Dep. Var. Mean	0.555	0.759	0.621	0.764	0.552	0.746
County FE	✓	✓	✓	✓	✓	✓
State-Year FE	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓
Observations	215,551	247,899	160,579	358,288	307,134	211,733

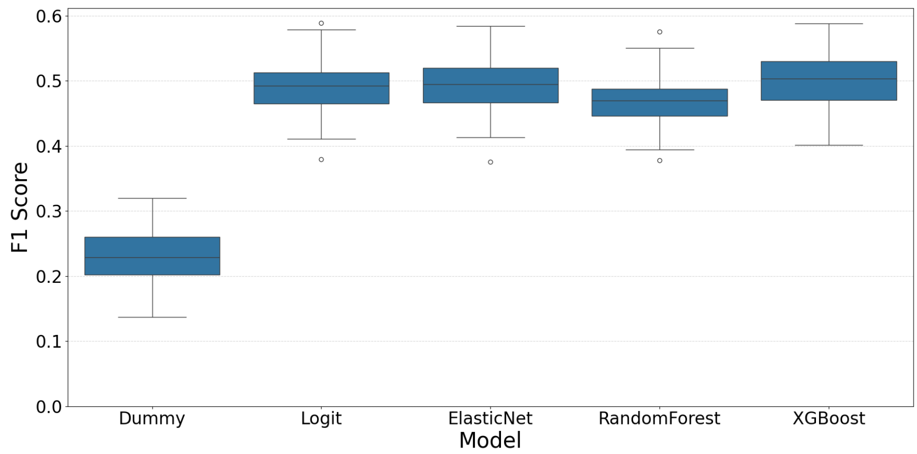
Notes: Estimates for the effect of high exposure to paywalls on political knowledge, estimated separately for different subgroups defined by yearly family income (Columns 1–2), highest attained educational degree (Columns 3-4), and age (Columns 5-6). The specification is motivated by Equation (3), estimated and aggregated using the Callaway-Sant'Anna estimator. The outcome is the share of correct answers by survey respondents to eight political knowledge questions (CES, 2006–2021), residualized on individual demographic controls – categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2) – as well as county and state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Table A6. Correlation of (continuous) paywall exposure and political knowledge - Heterogeneity by paying for news

Dependent Variable:	Share Correct Answers			
Pays for News (Pred.):	No		Yes	
	(1)	(2)	(3)	(4)
Paywall Exposure	-0.029**	-0.022**	-0.003	0.010
	(0.011)	(0.011)	(0.015)	(0.015)
Dep. Var. Mean	0.693	0.693	0.856	0.856
County FE	✓	✓	✓	✓
Year FE	✓		✓	
State-Year FE		✓		✓
Indiv. Demographics	✓	✓	✓	✓
Observations	413,797	413,797	105,070	105,070
R^2	0.297	0.307	0.285	0.308

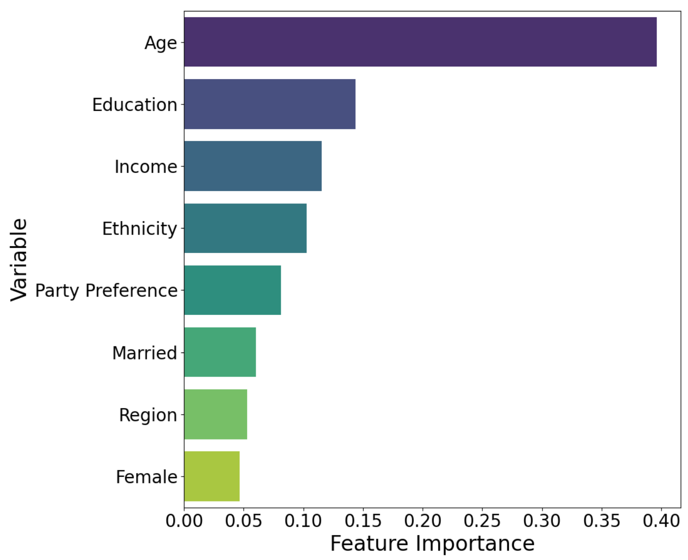
Notes: Estimates from regressing the share of correct answers to eight political knowledge questions on paywall exposure, separately by whether individuals are predicted to pay for a news subscription. All specifications control for fixed effects as indicated, as well as individual demographic controls comprised of categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2). Data are from the CES 2006–2021. Survey weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figure A7. Prediction of paying for news - Benchmark of prediction algorithms



Notes: Distribution of F1 scores from 100-fold cross-validation of the best estimator in each model category. The dummy classifier predicts a positive outcome at random, with probability equal to the share of positive cases in the training data. For the other classifiers, the best estimator was selected via successive halving random search over a manually defined starting grid, using 4-fold cross-validation. The target is an indicator for whether an individual reports paying for a news subscription, based on the Pew Research Center (2018) survey of 34,518 respondents. Explanatory variables include categorical indicators for age (4 groups), education (6), family income (9), ethnicity (2), party preference (4), marital status (2), region (4), and sex (2). Sample weights are applied during training.

Appendix Figure A8. Prediction of paying for news - Feature importances



Notes: Impurity-based feature importances for variable categories from the best XGBoost model predicting whether an individual pays for news, using data from the Pew Research Center (2018) survey. Explanatory variables include categorical indicators for age (4 groups), education (6), family income (9), ethnicity (2), party preference (4), marital status (2), region (4), and sex (2). Sample weights are applied during training.

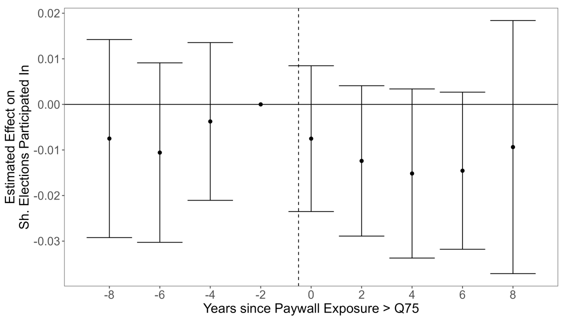
Appendix Table A7. Effect of paywalls on electoral participation - Index Decomposition (Individuals not Paying for News)

Dependent Variable:	Participation in Election						
Paying for News (Pred.):	No						
	President	House Rep.	Senators	Governor	General, any	State Rep.	State Sen.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Self-Reported							
Paywall Exposure > Q75	-0.0021	-0.0139**	-0.0097	-0.0225**	-0.0114**	-0.0153**	-0.0121*
	(0.0060)	(0.0060)	(0.0068)	(0.0090)	(0.0058)	(0.0062)	(0.0063)
Dep. Var. Mean	0.6460	0.5430	0.5920	0.5690	0.7030	0.5450	0.5360
County FE	✓	✓	✓	✓	✓	✓	✓
State-Year FE	✓	✓	✓	✓	✓	✓	✓
Individual Controls	✓	✓	✓	✓	✓	✓	✓
Observations	142,277	279,101	187,454	130,002	219,784	245,259	241,568
Panel B: Validated							
Paywall Exposure > Q75	-0.0048	-0.0135*	-0.0150**	-0.0051	-0.0117		
	(0.0066)	(0.0072)	(0.0075)	(0.0104)	(0.0072)		
Dep. Var. Mean	0.5450	0.4550	0.4980	0.4720	0.5270		
County FE	✓	✓	✓	✓	✓		
State-Year FE	✓	✓	✓	✓	✓		
Individual Controls	✓	✓	✓	✓	✓		
Observations	142,277	253,584	168,268	110,144	279,457		

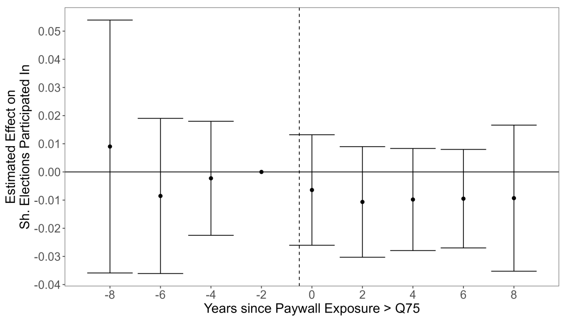
Notes: Estimates for the effect of high exposure to paywalls on electoral participation, among individuals predicted not to pay for a news subscription. The specification is motivated by Equation (3), estimated and aggregated using the Callaway-Sant'Anna estimator. he outcome is an indicator for participating in the election (CES, 2008–2020). Panel A uses self-reported voting; Panel B uses validated data from matching survey respondents to official records (Ansolabehere and Hersh, 2012). All specifications residualize the outcome on individual demographic controls – categorical indicators for sex (2 groups), age (5), ethnicity (2), education (6), employment status (6), family income (5), and children (2) – as well as county fixed effects and state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75$^{\text{th}}$ percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Appendix Figure A9. Effect of paywalls on participation in elections - Individuals not paying for news

(a) Self-reported

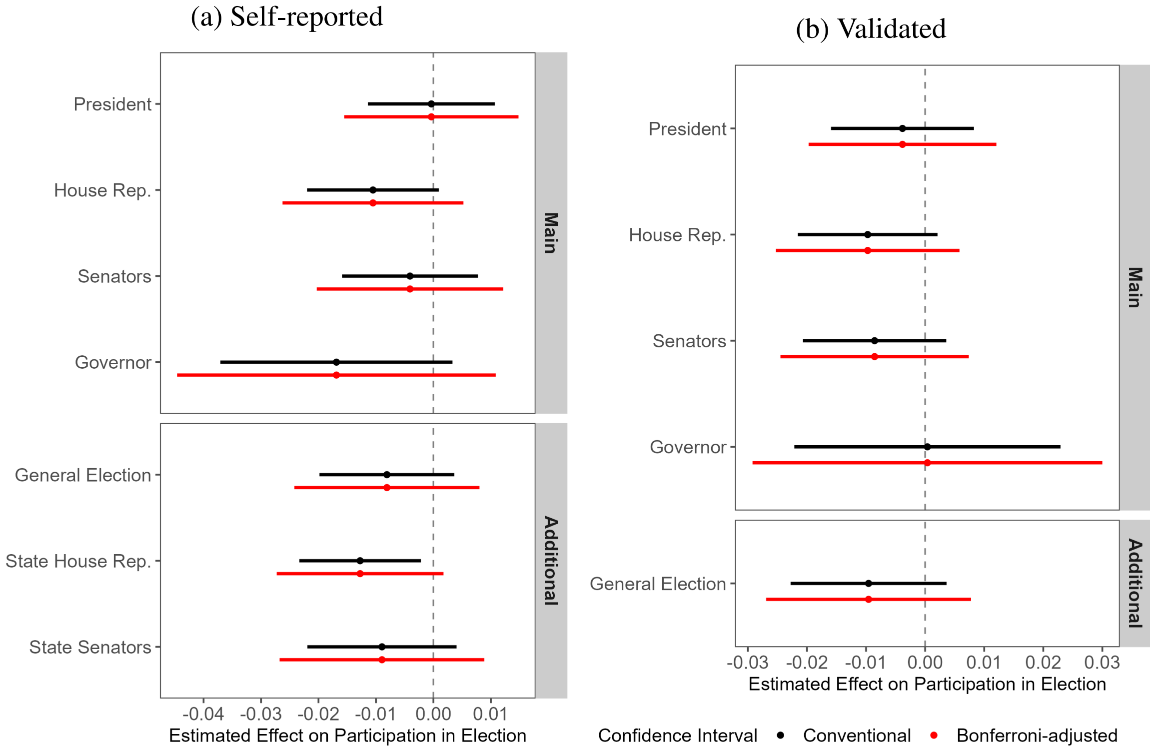


(b) Validated



Notes: Bi-yearly estimates for the effect of high exposure to paywalls on electoral participation, among individuals predicted not to pay for a news subscription. The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is the share of elections participated in among elections held for president, house representative, senators, and governor (CES, 2008–2020), residualized on individual demographic controls, county fixed effects, and state-by-year fixed effects. Panel (a) uses self-reported voting data, while Panel (b) uses validated data from matching survey respondents to official voting records (Ansolabehere and Hersh, 2012). Counties are treated once paywall exposure exceeds the 75th percentile. The control group consists of not-yet-treated and never-treated counties. Vertical bars show bootstrapped 90 percent uniform confidence intervals. Sample weights are included. Standard errors are clustered by county.

Appendix Figure A10. Effect of paywalls on participation in elections - Index decomposition
(Individuals not paying for news)



Notes: Each row represents an estimate for the effect of high exposure to paywalls on participating in the respective election, among individuals predicted not to pay for a news subscription. Panel (a) uses self-reported voting data, while Panel (b) uses validated data from matching survey respondents to official voting records (Ansolabehere and Hersh, 2012). In each panel, the top part refers to the elections included in the main index for participation, while the bottom part refers to other available data in the CES. Horizontal bars show bootstrapped 95 percent confidence intervals, estimated conventionally (black) or with Bonferroni adjustment (red). The specification is motivated by Equation (3) and estimated using the Callaway-Sant'Anna estimator. The outcome is an indicator for participating in the election (CES, 2008–2020), residualized on individual demographic controls, county fixed effects, and state-by-year fixed effects. Counties are treated once paywall exposure exceeds the 75th percentile. The control group consists of not-yet-treated and never-treated counties. Sample weights are included. Standard errors are clustered by county.