Political Effects of Newspaper Paywalls

Julian Streyczek
Università Bocconi, Milan, Italy
Department of Economics

August 26^{th}, 2025

Motivation

- **Information** for voters largely provided by **private actors**

  $\Rightarrow$ Shocks to news markets may affect **democratic outcomes**

  (e.g., Besley and Burgess 2002; Gentzkow 2006; DellaVigna and Kaplan 2007; Snyder Jr and Strömberg 2010; Falck et al. 2014; Gavazza et al. 2019; Zhuravskaya et al. 2020)

- Recent example: **Paywalls**:
  - 2000s: Newspapers put most content on website for free
  - 2010s: Newspapers start charging price for full access to website
  - 2020: $>$ 90% of Top 100 US newspapers have paywall

Motivation

- **Information** for voters largely provided by **private actors**

  $\Rightarrow$ Shocks to news markets may affect **democratic outcomes**

  (e.g., Besley and Burgess 2002; Gentzkow 2006; DellaVigna and Kaplan 2007; Snyder Jr and Strömberg 2010; Falck et al. 2014; Gavazza et al. 2019; Zhuravskaya et al. 2020)

- Recent example: **Paywalls**:
  - 2000s: Newspapers put most content on website for free
  - 2010s: Newspapers start charging price for full access to website
  - 2020: $>$ 90% of Top 100 US newspapers have paywall

$\Rightarrow$ **Paywalls increase cost of obtaining information**

Did the introduction of paywalls on US newspapers websites affect political outcomes?

This paper

- **Staggered introduction of paywalls** across US newspapers and regions
- Measure **counties' exposure to paywalls** over time: Use print circulation to proxy which regions affected by paywalls
- Main outcomes from **large-scale survey** *(Cooperative Election Study)*:
  - Political knowledge (representatives, majorities)
  - Electoral participation
- **Contribution:** First evidence on political effects of paywalls

Roadmap

1. **Data**
2. **Effect on news consumption**
3. **Effect on knowledge**
4. **Effect on voting turnout**

Data

Data (1/3)

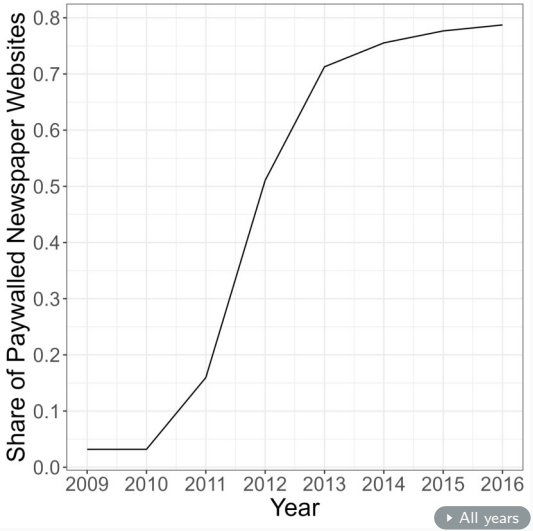
1. **Newspapers**
   - **Sample:** Top 100 US daily newspapers by print circulation in 2011
   - **Print circulation** from *Alliance for Audited Media*:
     - Number of print copies sold per newspaper and county in 2011 (301 newspapers)
   - **Page views** from *Alexa Internet*: (for Top 80)
     - Collected via *Alexa* browser toolbar (no mobile, no tablet)
     - Number of page views (clicks) to domain per million users
     - Daily, 2010–2017, globally (no user-level data)
   - **Characteristics:**
     - Mostly local, 3 National *(New York Times, USA Today, Wall Street Journal)*
     - No exits
     - 76% of total US print circulation in 2011 (based on AAM sample)
     - 1.5x the combined page views for ABC, CBS, NBC, Fox, AP, Reuters in 2010

Data (2/3)

2. Paywalls on newspaper websites

- Introduction dates, hand-collected (news coverage, website snapshots)
- Common characteristics (early 2010s):
  - Metered: *Any* first 5-20 clicks per month free
  - Leaky: Could circumvent with sufficient effort
- Here: Focus on *existence* of paywalls, abstract from *features*

▸ Subscription prices



▸ All years

Data (3/3)

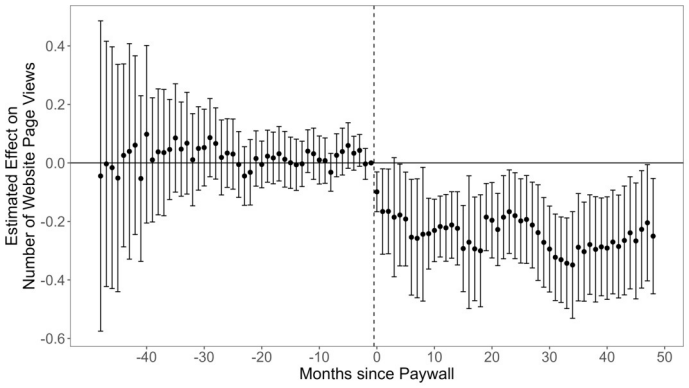
3. Survey: Cooperative Election Study (CES)

- 10-60k respondents per year, 2006–2021, by county, repeated cross section
- **Political Knowledge:** Questions about political representatives & party majorities
- **Electoral Participation:** Voting in elections: President, House, Senate, Governor (self-reported + validated)

Effects on online news consumption

Website page views decrease by 25-30% after paywall introduction

$$\log(\textit{Pageviews}_{n,t}) = \sum_{\tau \neq -1} \beta_\tau \, D^{\tau}_{n,t} + \textit{NewspaperFE}_n + \textit{YearMonthFE}_t + \delta' \mathbf{X_{n,t}} + \varepsilon_{n,t}$$



- TE $= -0.280^{***}$
- Not driven by selection or spillovers
- Substitution to print is rare (Pattabhiramaiah et al. 2019)
- Timing: Topic changes unlikely main driver

Estimated via Callaway-Sant'Anna. 55 Treated, 17 Control, 2010–2017, $N = 6,912$.
Controls x Year-Month: Population density, income, partisanship.
SE clustered by newspaper. Bars denote 95 percent confidence intervals.

▸ Sample ▸ Identification ▸ Control variables ▸ Estimators: Figure ▸ Estimators: Table ▸ Control group ▸ Spillovers

Effects on political knowledge

Outcome: Political Knowledge

- 8 knowledge questions:
  (based on respondent's place of residence)

  1. **Which party does [name of political representative] represent?**
     (4 Questions: Governor, House Representative, 2 Senators)
  2. **Which party has the majority in [legislative body]?**
     (4 Questions: House, Senate, State House, State Senate)

  ▸ Distribution ▸ By Question ▸ Trends ▸ Map

- **Compile into index:** Share of correctly answered questions
  (Robustness: 1^{st} Principal Component)

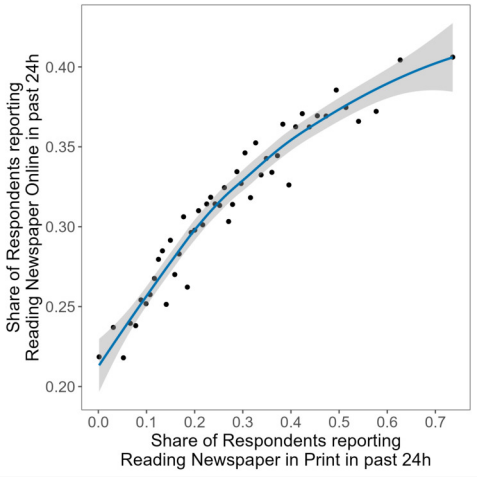
Treatment: Idea

Data challenge:

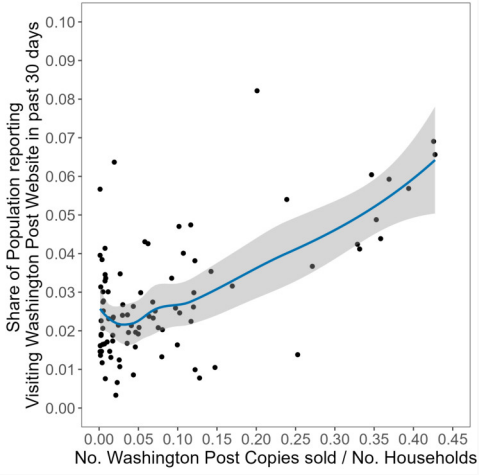
- **Goal:** Measure county-level exposure to paywalls over time
- **Ideal:** Measure *online* news consumption per newspaper on county-level
  $\Rightarrow$ Unavailable
- **Alternative:** Use print circulation as proxy for online readership
- **Intuition**: If print version of newspaper is relevant in county, then also online version is relevant
  (Most newspapers in sample are *local*)

Treatment: County-level print circulation predicts online readership

CES: More offline readers predicts more online readers



Washington Post: Higher print circulation predicts more website visits



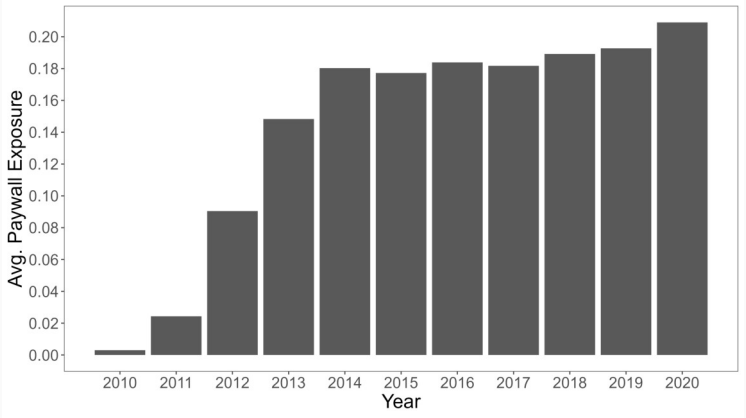
Treatment: County-level exposure to paywalls

$$\begin{aligned} \textit{Paywall Exposure}_{c,t} = & \text{ Per household circulation of newspapers} \\ & \text{ that have been affected by a paywall} \\ & \text{ by year } t \\ = & \sum_{n \in \textit{Newspapers}} \frac{\textit{Circulation}^{n}_{c,2010} \times I(\textit{Paywall}^{n}_{t})}{\textit{Households}_{c,2010}} \end{aligned}$$

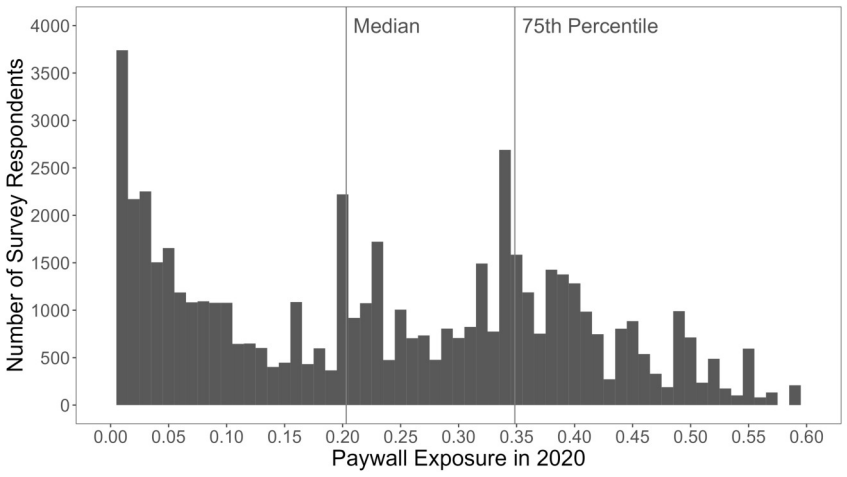
- $\textit{Circulation}^{n}_{c,2010}$: Circulation of newspaper n (in county c in 2010)
- $\textit{Households}_{c,2010}$: Number of households (in county c in 2010)
- $I(\textit{Paywall}^{n}_{t})$: Paywall indicator =1 if newspaper n has paywall in year t

$\Rightarrow$ **Proxy** for share of households affected by paywall

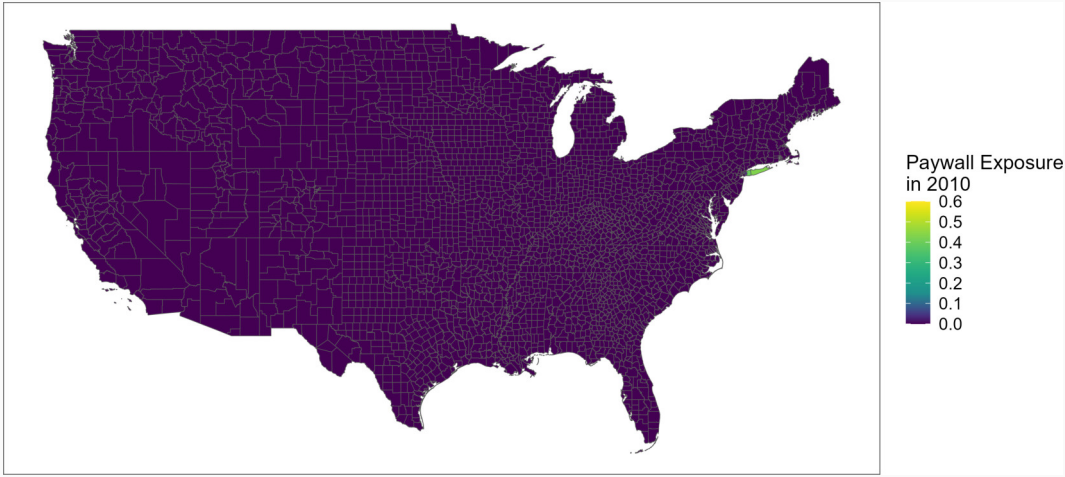
Paywall Exposure increases between 2010 and 2014



Large variation in Paywall Exposure in 2020

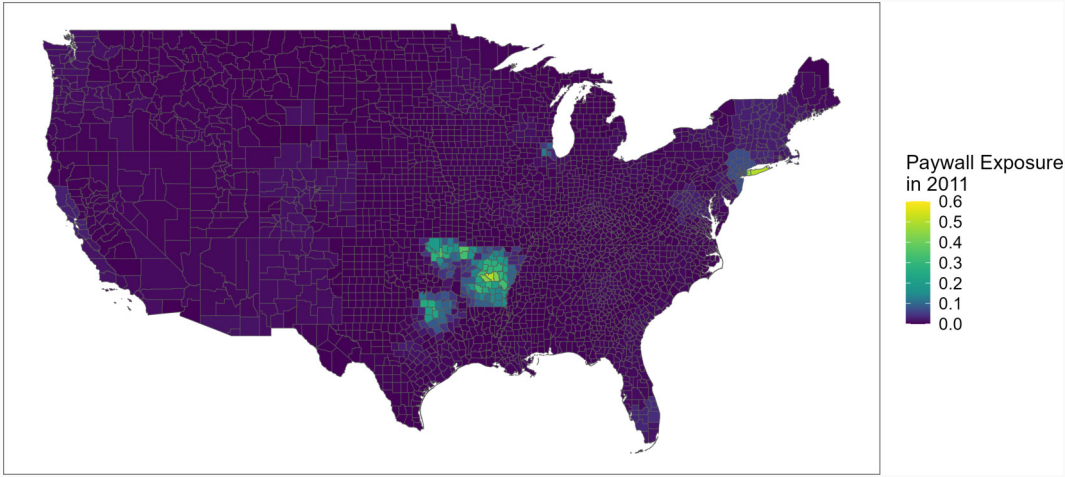


Paywalls affected all regions of the US (2010)



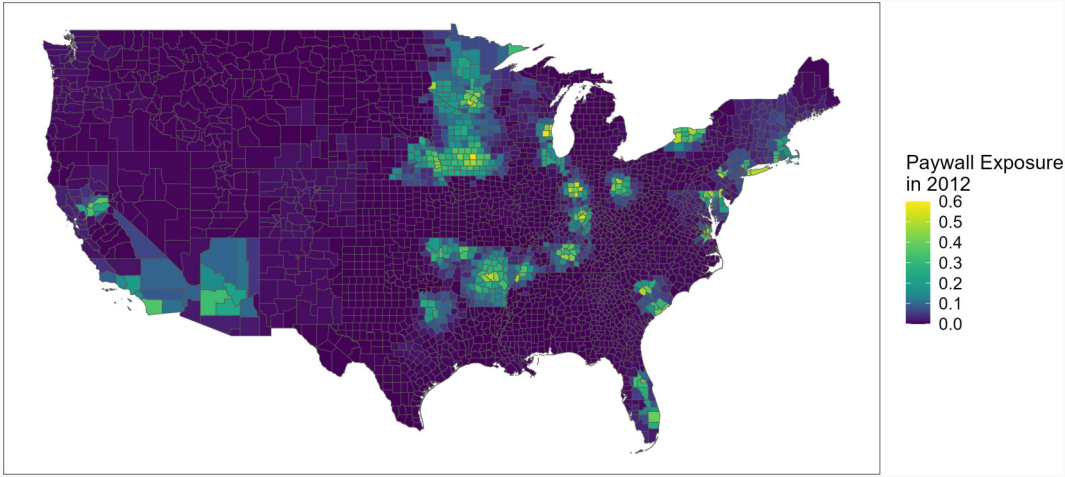
▸ Residualized by DMA-Year ▸ Correlation with observables

Paywalls affected all regions of the US (2011)



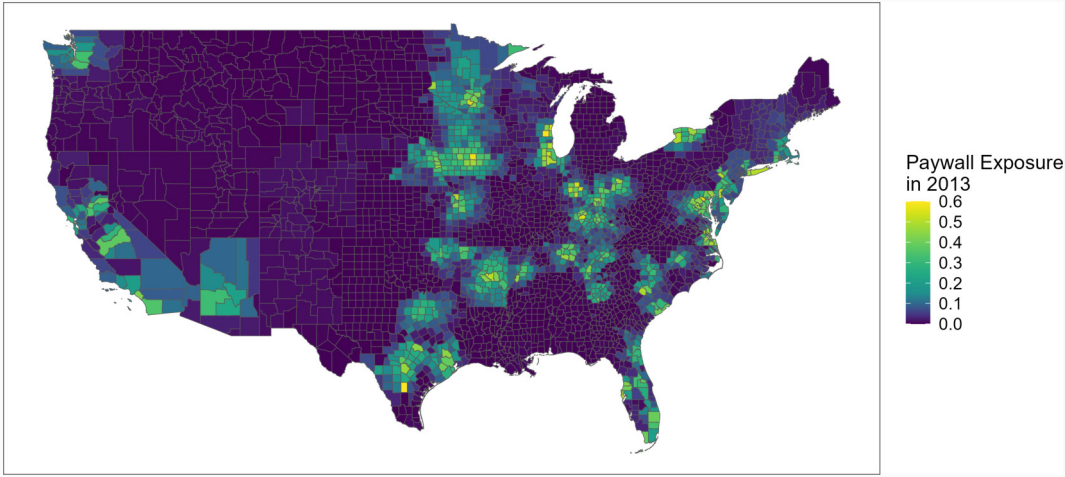
▸ Residualized by State-Year ▸ Correlation with observables

Paywalls affected all regions of the US (2012)



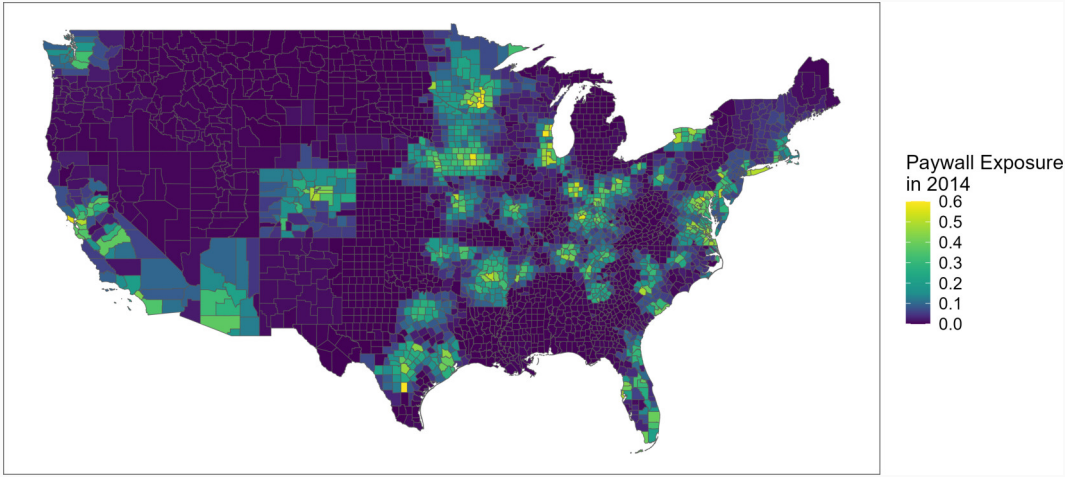
▸ Residualized by State-Year ▸ Correlation with observables

Paywalls affected all regions of the US (2013)



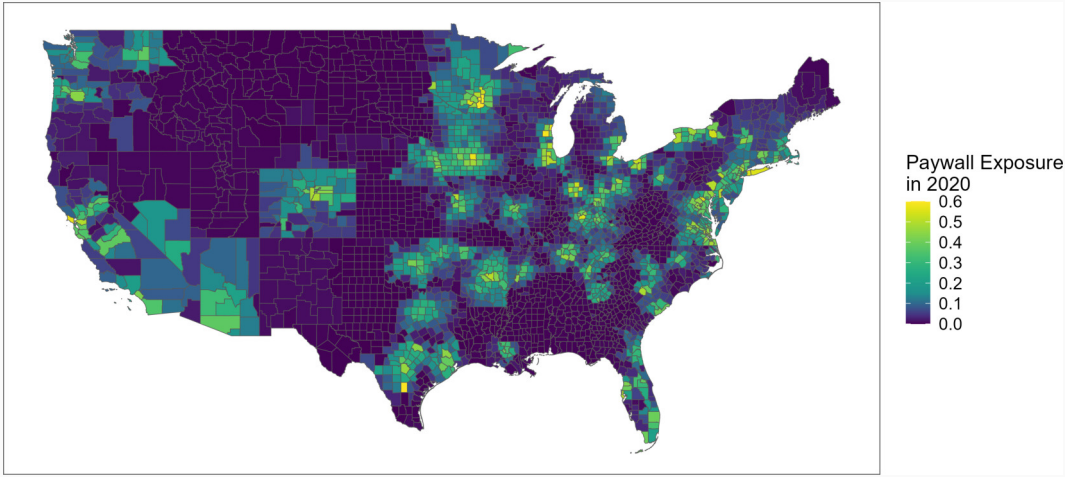
▸ Residualized by State-Year ▸ Correlation with observables

Paywalls affected all regions of the US (2014)



▸ Residualized by State-Year ▸ Correlation with observables

Paywalls affected all regions of the US (2020)



▸ Residualized by State-Year ▸ Correlation with observables

Empirical strategy

Staggered diff-in-diff:

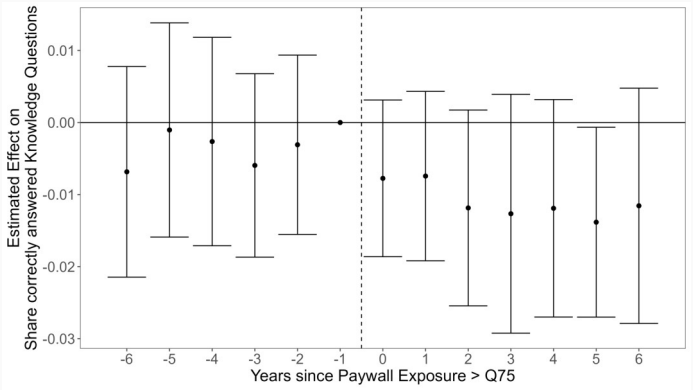
$$Knowledge_{i,c,t} = \sum_{\tau \neq -1} \beta_\tau D^\tau_{c,t} + \alpha_c + \gamma_{s(c),t} + \delta' \mathbf{X}_{i,c,t} + \varepsilon_{i,c,t}$$

- $Knowledge_{i,c,t}$: Share correct answers by survey respondent i in county c (state s) in year t
- $D^\tau_{c,t}$: Years since paywall exposure in county c first exceeds 75$^{\text{th}}$ percentile (end of sample) (See Guriev et al. 2021, QJE)
- $\mathbf{X}_{i,c,t}$: Individual controls
  (sex (2), age (5), ethnicity (2), education (6), employment (6), family income (5), children (2))
- **Robust estimator:** Callaway-Sant'Anna (Residualize on controls and FE first)

Identification

- **Identifying assumption:** Parallel trends of counties with high and low paywall exposure in absence of paywalls
- **Challenge:** Omitted variables correlated with both paywalls and knowledge / Measurement error in Paywall Exposure
  - $\Rightarrow$ Staggeredness
  - $\Rightarrow$ Individual characteristics x Year-FE
- **Other challenges:**
  - Correlation with regional politics ($\Rightarrow$ State-Year FE)
  - Heterogeneous effects ($\Rightarrow$ Robust estimator)
  - Selection of Newspapers ($\Rightarrow$ Sensitivity checks)
- **Remaining threats:** Variation within State-Year that is correlated with knowledge, but uncorrelated with individual characteristics

Paywalls reduce share correctly answered political knowledge questions by 1.6%



Estimated via Callaway-Sant'Anna. 2006–2021, $N = 518,867$
SE clustered by county. Bars denote 90 percent uniform confidence intervals.

▸ Control variables ▸ Continuous treatment ▸ Bins
▸ Thresholds ▸ Index decomposition ▸ Heterogeneity ▸ "Persuasion rate"

- TE $= -0.012^{**}$ (Baseline: 0.745)
- Interpretation:
  i) 1 more incorrect answer per 8 respondents
  ii) 1 in 46 respondents loses all knowledge
- Decreasing knowledge across all questions
- Driven by lower income and lower formal education

Decrease in knowledge entirely by individuals not paying for news

Estimated via Callaway-Sant'Anna. 2006–2021, $N = 518,867$
SE clustered by county. Bars denote 90 percent uniform confidence intervals.

▸ Prediction Performance ▸ Feature Importances ▸ External validity ▸ Control variables

- Individuals predicted to pay / not pay for news, based on large survey (2018 Pew-Ipsos Local News Survey, N=34,897)
- TE, not paying $= -0.015^{**}$
- TE, paying $= 0.002$

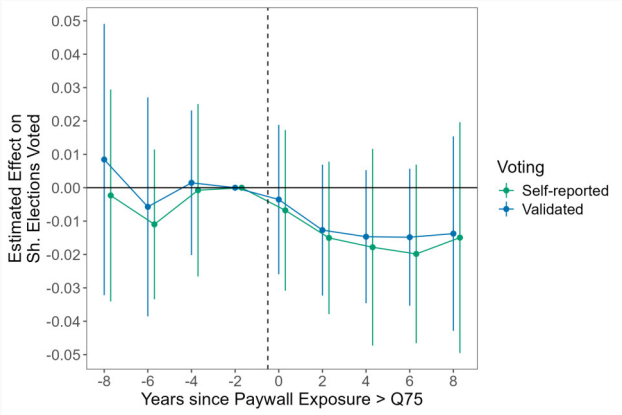
Effects on voting turnout

Outcome: Electoral participation

- **Participation** in 4 types of elections,
  both self-reported and validated:
  - President
  - House
  - Senate
  - Governor

- **Compile into index:** Share of elections voted in, among all held in year
  (Note: State-Year FE account for set of available elections)

Paywalls reduce voting in elections, among individuals unlikely to subscribe, in less partisan counties



Estimated via Callaway-Sant'Anna. 2008–2020, $N = 300,286$
SE clustered by county. Bars denote 90 percent confidence intervals.

▸ Table: Self-reported ▸ Table: Validated ▸ Index Decomposition

- TE, self-reported $= -0.0147^{***}$
- TE, validated $= -0.0115^{*}$
- Driven by all types of elections
- Full Sample: Less robust but qualitatively similar

Conclusion

Conclusion

- **Paywalls displace readers** from major regional newspaper websites
- **Declines in political knowledge** and **electoral turnout**
- Consistent with switch to news sources with less intensive coverage of politics

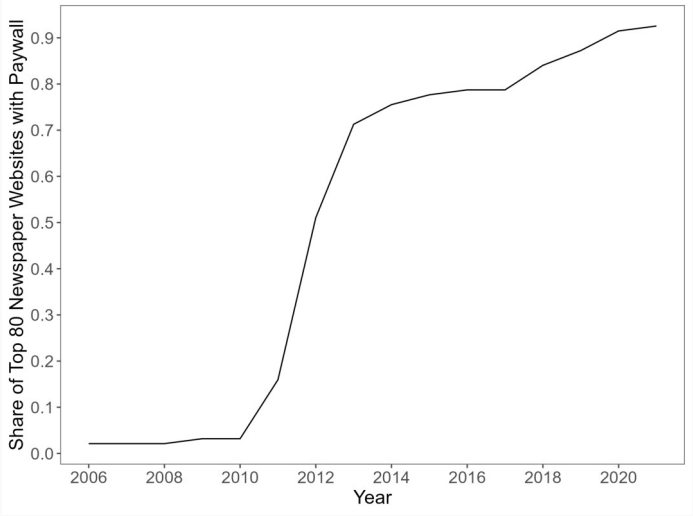
  $\Rightarrow$ Easy and equitable **access to information** is important for **democracy**

I'm looking forward to hearing your thoughts!

julian.streyczek@phd.unibocconi.it

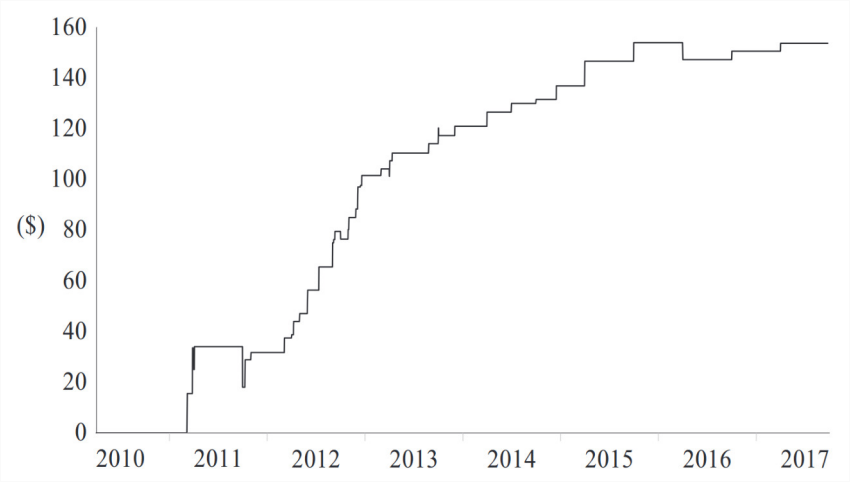
Appendix

Many paywalls introduced between 2010 and 2014



Back

Price for yearly digital subscriptions to Top 42 US newspapers



Source: Kim et al. (2020)

Back

Effect on pageviews: Sample

- **Sample:** 72 newspaper websites
- **Period:** January 2010 – December 2017, monthly
- **Treated:** 55 newspapers with paywall introduced in sample period
- **Control:** 2 earlier, 11 later, 4 never

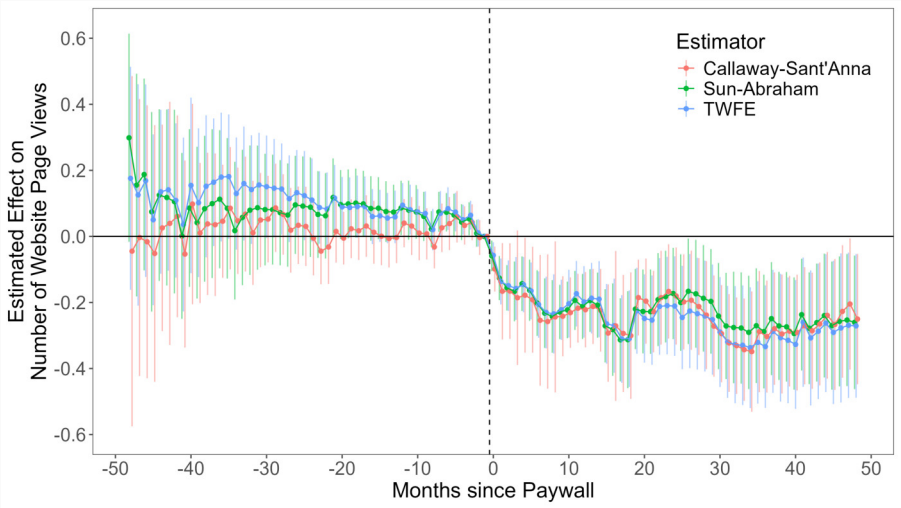
Back

Identification

- **Identifying assumption:** Parallel trends between paywalled and non-paywalled page views in absence of paywall
- **Challenge 1: Selection**
  - ⇒ Pre-trends
  - ⇒ Time-varying audience controls
  - ⇒ Vary control group composition
- **Challenge 2: Spillovers**
  - ⇒ Regional segmentation
  - ⇒ Robustness against excluding largest paywalled / smallest control newspapers
  - ⇒ To do: Stacked event study, Local projection

Back

Results are robust to alternative estimators



Back

Page views decrease by 25-30% after paywall introduction

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

Robust Callaway-Sant'Anna estimator. SE clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Back

Results are robust to alternative 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		

Robust Callaway-Sant'Anna estimator. SE clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Back

Results are robust to definition of control group

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

Robust Callaway-Sant'Anna estimator. SE clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

▸ Back

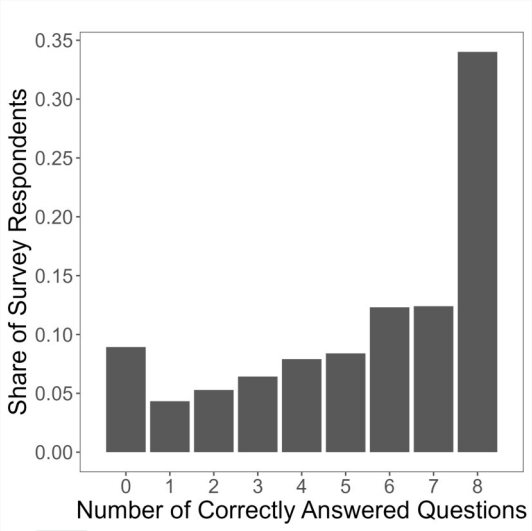
Results are robust to excluding national + largest newspapers

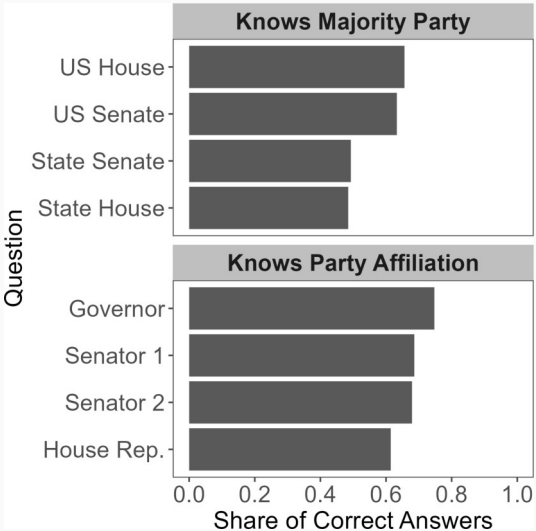
Dependent Var.:	Log(Pageviews)				
	Exclude Newspapers				Add Weights
	National	Top 15	Top 15 Paywalled	Bottom 5 Control	Weights
	(1)	(2)	(3)	(4)	(5)
Paywall	-0.282***	-0.332***	-0.357***	-0.320***	-0.273***
	(0.065)	(0.071)	(0.073)	(0.075)	(0.068)
Newspaper FE	✓	✓	✓	✓	✓
Month-Year FE	✓	✓	✓	✓	✓
Market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

Robust Callaway-Sant'Anna estimator. SE clustered by newspaper. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Back

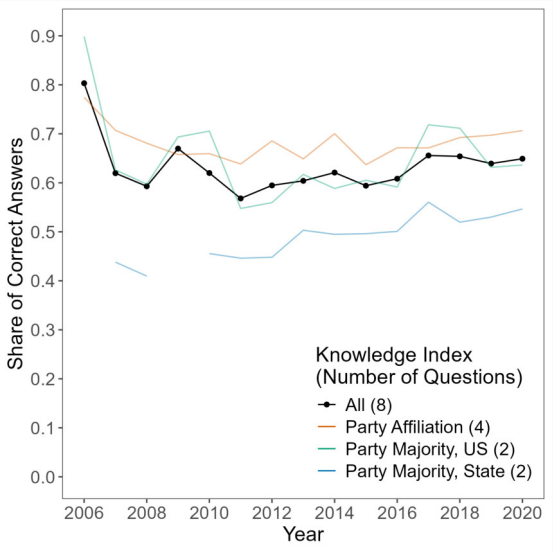
Distribution of Answers to Political Knowledge Questions





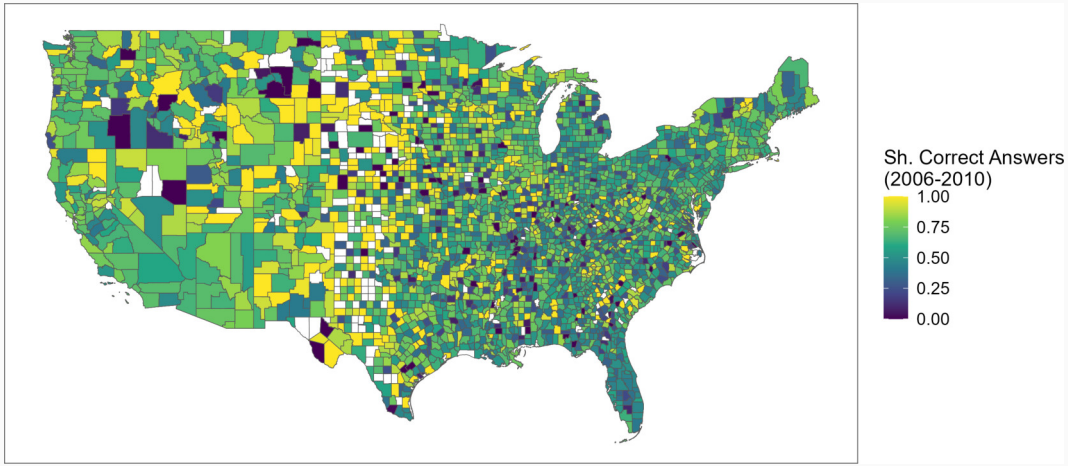
Back

Number Correctly Answered Political Knowledge Questions by Year



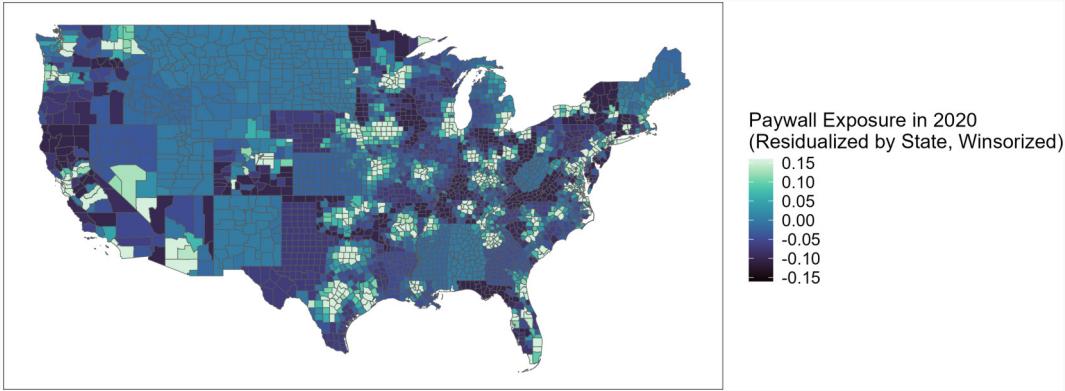
Back

Sh. Correctly Answered Political Knowledge Questions by County



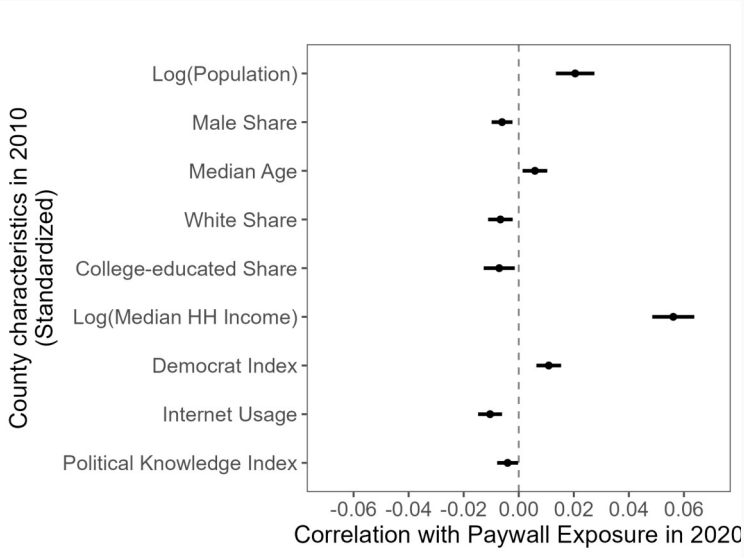
Back

Paywall Exposure by county in 2020, residualized by State-Year



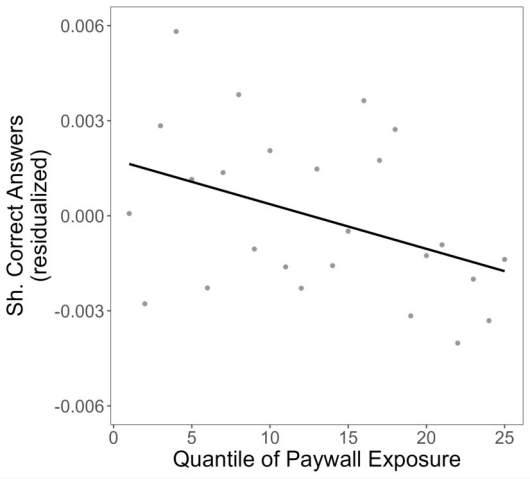
Back: 2020

County variables in 2010 correlated with Paywall Exposure in 2020



Back: Map 2020

Paywalls reduce online news consumption



Back

Paywalls reduce online news consumption

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

SE clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

▸ Back

Paywalls reduce online news consumption

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

SE clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Back

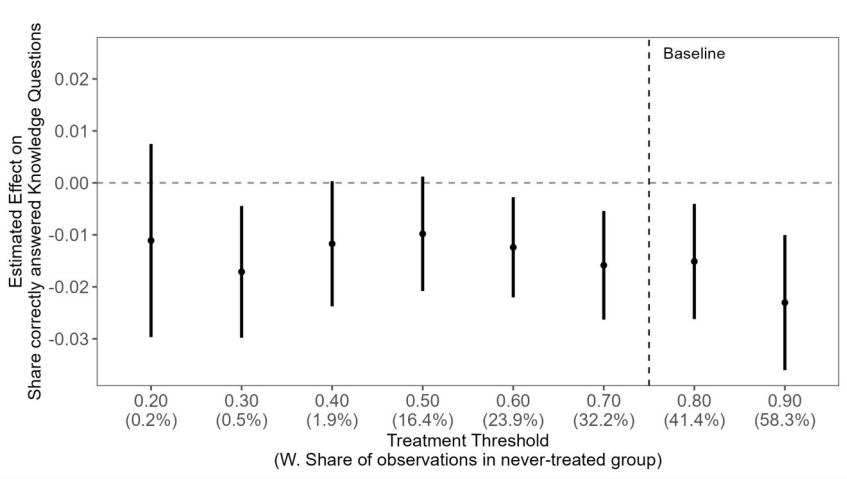
Paywalls reduce share of correctly answered political knowledge questions

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: Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

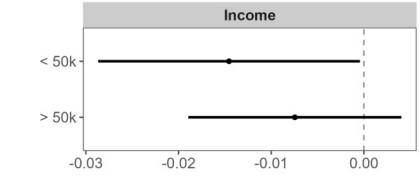
Back

Results are robust to choice of threshold



Back

Driven by individuals with lower income and education



▸ Back to results

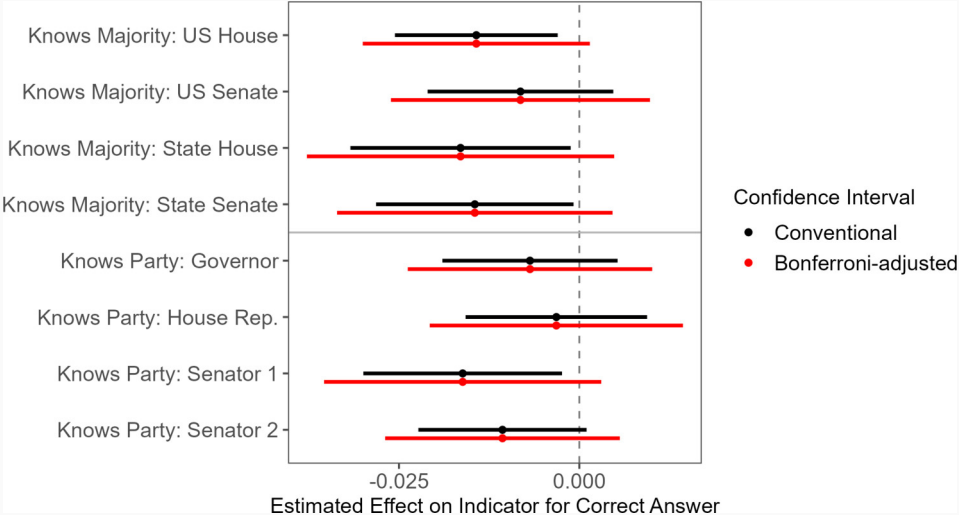
Driven by individuals with lower income and education

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: Callaway-Sant'Anna estimator. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

▸ Back

Knowledge decreases for both federal- and state-level politics



Back to results

"Persuasion rate" is moderate

Definition: Share of individuals who switched from correct to incorrect answer because of paywalls, among those exposed to paywalls:

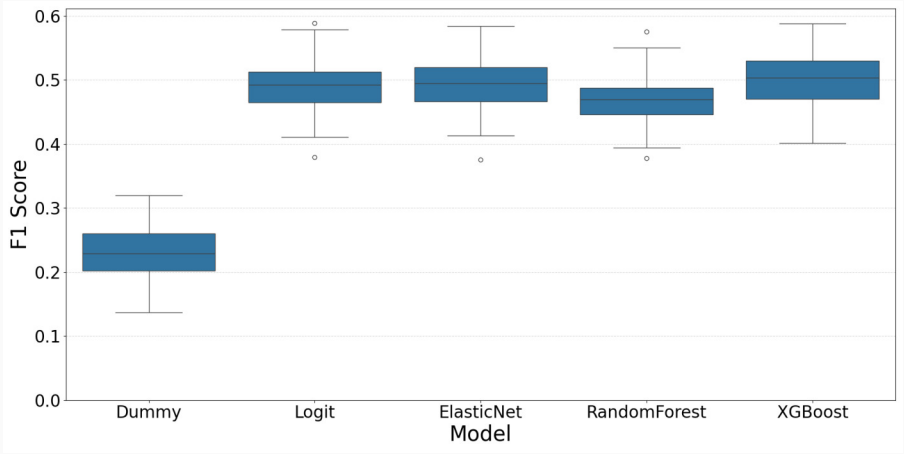
Back-of-envelope calculation: (DellaVigna and Gentzkow, 2010)

$$\begin{aligned} f &= \frac{y_T - y_C}{e_T - e_C} \frac{1}{1 - y_0} \\ &= \frac{0.012}{0.5 * 0.5} \frac{1}{0.745} \\ &= 0.048 \end{aligned}$$

- $y_T - y_C$: Effect of paywalls on knowledge
- $e_T - e_C$: $0.5 \approx$ (Share US adults reading news online 2010–2017)
  $\times\ 0.5 \approx$ (Differential paywall exposure between treated and control)
- $1 - y_0$: Baseline knowledge

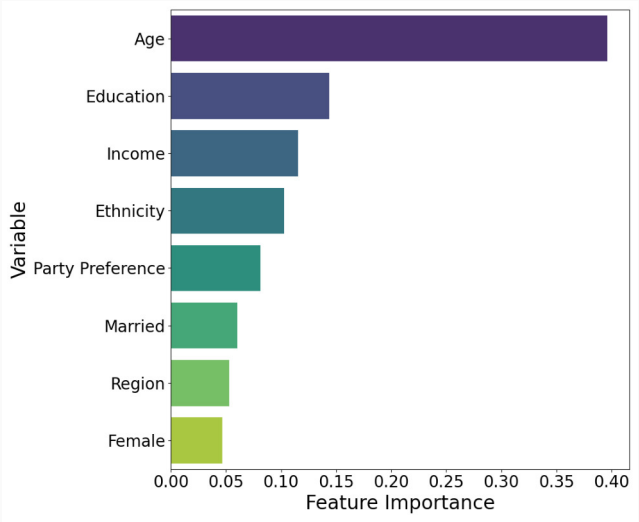
Back

Paying for News: Prediction Performance



Back

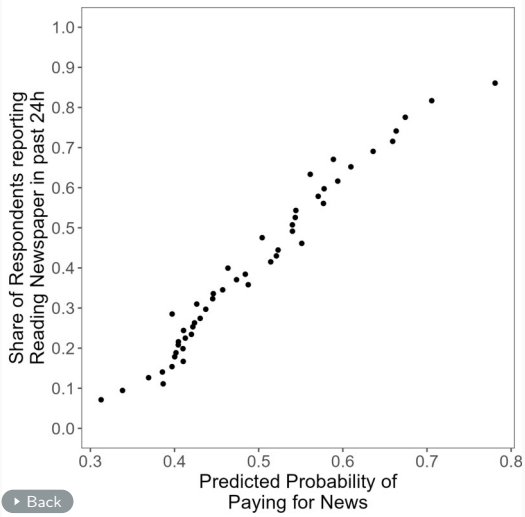
Paying for News: Feature Importance



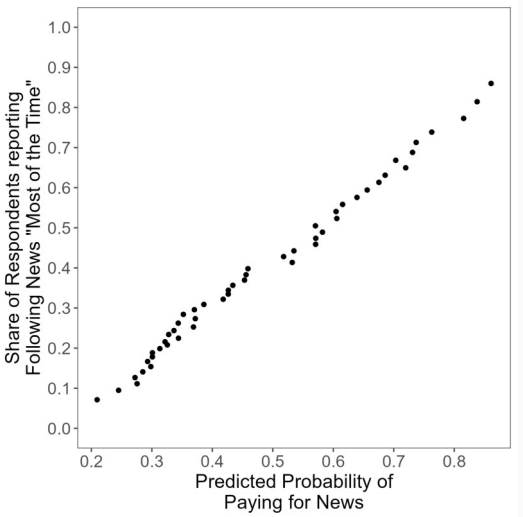
Back

Paying for News: External Validity

(a) Read newspaper in past 24h



(b) Follow news "most of the time"



Back

Decrease in knowledge entirely by individuals not 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: Callaway-Sant'Anna estimator. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Back

Paywalls reduces voting in elections among individuals unlikely to subscribe

Dependent Variable:	Participation in Elections, self-reported			
Paying for News (Pred.):	All		No	
County Partisanship Index:	All			0.3-0.7
	(1)	(2)	(3)	(4)
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

Notes: Callaway-Sant'Anna estimator. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Back

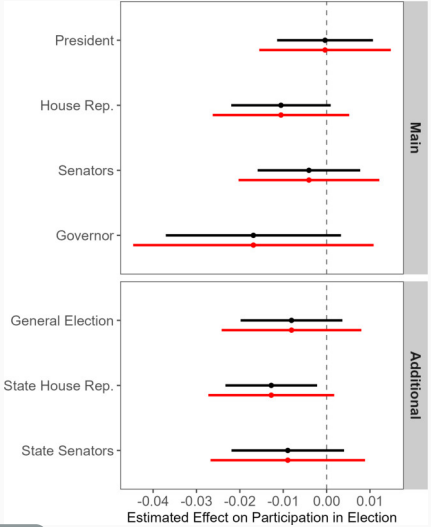
Paywalls reduce voting in elections among individuals unlikely to subscribe

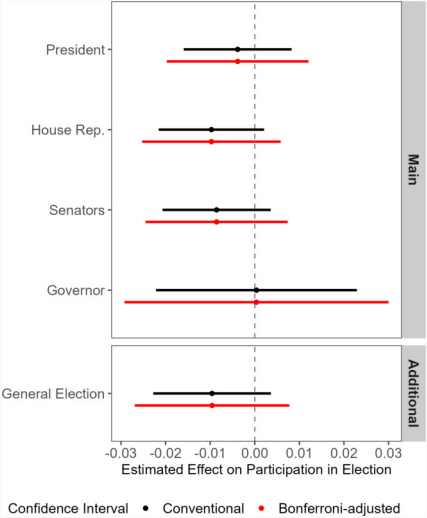
Dependent Variable:	Participation in Elections, validated			
Paying for News (Pred.):	All		No	
County Partisanship Index:	All			0.3-0.7
	(1)	(2)	(3)	(4)
Paywall Exposure > Q75	-0.0142**	-0.0055	-0.0090	-0.0115*
	(0.0064)	(0.0051)	(0.0057)	(0.0063)
Dep. Var. Mean	0.5089	0.5089	0.4798	0.4814
County FE	✓	✓	✓	✓
Year FE	✓			
State-Year FE		✓	✓	✓
Individual Controls	✓	✓	✓	✓
Observations	383,040	383,040	300,286	253,940

Notes: Callaway-Sant'Anna estimator. Sample weights are included. Standard errors (in parentheses) are clustered by county. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Back

Turnout decreases for federal and state elections





Back