

Election Coverage and Slant in Television News*

Gregory J. Martin[†] Ali Yurukoglu[‡]

PRELIMINARY AND INCOMPLETE
Please do not circulate without permission.

Abstract

This paper’s goal is to compare equilibrium news coverage during the 2012 US Presidential campaign to a benchmark of socially optimal news coverage. We specify a model where viewer-voters have utility for news stories driven by three considerations: (1) learning information that is relevant for the Presidential election, (2) consuming political news that matches their own ideology, and (3) consuming news for pure leisure. News channels choose topic coverage to maximize viewership. We calibrate the model to match high frequency data on individual level viewership and topic coverage by news channels, as well as lower frequency polling data.

*We thank Ellen Dudar and Bill Feininger of FourthWall Media for access to FourthWall’s viewership data, Alex Coblin for excellent research assistance, and seminar participants at Princeton, Yale, and the University of Lausanne Economics of Media Bias Workshop.

[†]Emory University.

[‡]Stanford Graduate School of Business and NBER.

News media play an essential role in the functioning of electoral institutions in democratic societies. Media outlets’ editorial decisions shape the information set to which voters have access when choosing between candidates, a responsibility that is consequential for vote choices even if voters are fully sophisticated Bayesians (Kamenica and Gentzkow, 2011). Without voter access to accurate information about candidates’ platforms and past performance in office, neither the preference aggregation nor the accountability functions of elections can be expected to perform well.

Yet, despite the instrumental importance of information in election outcomes, instrumental demand for information is not the only or even the primary force shaping the provision of politics news coverage. Indeed, in a large election no individual voter has more than an infinitesimal chance of being pivotal, meaning that the incentive for individuals to acquire information in order to make a better voting decision is very weak. Instead, non-instrumental sources of demand are likely to dominate. Viewers may consume political content as entertainment, valuing electoral politics’ ability to deliver exciting “suspense and surprise” (Ely et al., 2015) over the course of a campaign. Or they may consume news with an agreeable partisan slant for the psychological benefit of having their pre-existing beliefs confirmed (Mullainathan and Shleifer, 2005). News outlets, as profit-seeking businesses dependent on advertising and subscription revenue for survival, must cater to these tastes if they hope to remain viable.

This paper uses high-frequency household-level panel data on news consumption - specifically, cable and national network television news viewing - to decompose demand for news content into instrumental and non-instrumental components. We join this viewership data with an extensive data set on news provision generated from transcripts of television news programs. Our analysis makes use of two sources of variation in the instrumental value of information in our data. First, information farther in advance of the election date is less instrumentally valuable than information closer to the election date, because there is less time for the political situation to change in the interim. Once the election is over, this time

trend ends abruptly, as there is no longer any use in acquiring political information for the purpose of improving one’s vote choice. Second, cross-sectional variation across households is informative because it is only those voters for whom information revealed during the campaign might plausibly change their vote - the mythical “swing voter” - for whom political news has any instrumental use.

We build and estimate the parameters of a model of demand for news coverage that includes viewer tastes for information driven by all three mechanisms - the instrumental vote-choice improving value, the entertainment value of surprising new developments, and the prior-confirming affirmation of agreeable slant. News channels in the model select stories to report to maximize viewership given viewer tastes. The unpredictable arrival of breaking news events over time generates exogenous temporal variation in viewers’ preferences over news topics and thereby in channels’ coverage decisions, allowing us to separately identify the components of viewers’ utility.

With estimates of the parameters in hand, we can use the model to understand how market forces shape the coverage that viewers get from TV news, relative to the coverage that would be provided by a social planner seeking to maximize the quality of viewers’ information at election time. In this sense, the model allows us to measure the direction and magnitude of informational externalities generated by non-instrumental tastes for information. Ely et al. (2015) suggest the existence of positive externalities, noting that “despite this lack of a direct [instrumental] incentive, many voters do in fact follow political news and watch political debates, thus becoming an informed electorate (p. 216).” But negative externalities are also possible, if media outlets focus political coverage on topics - such as campaign gaffes or sex scandals - that are surprising and entertaining but contain relatively little information on candidates’ policy goals or performance in office.

Our analysis connects to an existing empirical and theoretical literature on media effects in campaigns and electoral politics. Prat (2017) shows that in the US, the media ownership groups with the most dedicated audiences are the conglomerates that own cable news

channels. These conglomerates have, in Prat’s term, significant “media power:” the ability, through selective presentation of information, to engineer an election victory for a candidate that would otherwise lose. Strömberg (2004, 2001), models the choice of content provision by media outlets seeking to maximize readership in a static setting, finding that news coverage is tailored to consumers with high private value of news. Our model has an analogous property, but adds some more subtle dynamic implications relevant to our election-season empirical application. Gentzkow and Shapiro (2010) consider the determinants of slant in local newspapers’ political coverage, distinguishing owner-driven from reader-driven variation in slant. DellaVigna and Kaplan (2007), Enikolopov et al. (2011), Durante and Knight (2012), Peisakhin and Rozenas (2017), and Martin and Yurukoglu (2017) focus, like our paper, on political news on TV, although they focus on the persuasive effects of partisan media outlets as opposed to our emphasis on informational externalities of tastes for politics news as entertainment. Garcia-Jimeno and Yildirim (2017) consider dynamic interactions between media and candidates over the course of a campaign. Their model emphasizes candidates’ incentives to reveal or conceal information to media; we treat the arrival of stories to news outlets as exogenous and focus on news outlets’ choice of what stories to cover, and viewers’ choices of what stories to watch.

1 Data

This paper focuses on national television news broadcasts, including the three cable news networks CNN, the Fox News Channel (FNC) and MSNBC and the national evening news programs on ABC, CBS, NBC and PBS. Despite the proliferation of online news sources, social media, and the like, television news maintains a large and devoted following. In the 2016 election, according to Pew Media Research, 38% of a nationally representative sample of American voters named one of the three cable news channels as their “main source” for news about the campaign, and an additional 15% named one of the national network broadcasters

(Gottfried et al., 2017).

We employ high-frequency panel data which allows for precise measurement of viewer tastes. We match this data to information on channels’ topical coverage and political slant derived from transcripts of news show broadcasts, allowing measurement of viewer reaction to variation along these dimensions. Our data is disaggregated to the household level, allowing us to measure differential responses across political and demographic dimensions. As a result, we can estimate effects on both the size and the composition of the audience that result from changes in channels’ coverage decisions.

We use four primary data sets in our analyses: household-level set-top-box (STB) data, aggregate television ratings data, cable news show transcripts, and daily presidential poll results. The high-frequency, household-level STB data allow us to fairly precisely measure viewers’ reactions to the content provided by the news channels, which we measure using the database of news show transcripts. Household-level demographics associated with the STB data allow us to estimate differential responses by viewer types, including along political dimensions. We use the aggregate ratings data both to validate the patterns evident in the STB data, and to adjust for the non-nationally-representative set of markets included in our STB sample. Finally, polling data from nationally representative polls help to identify the timing of important events during the campaign. We describe each of these data sets in turn.

STB Data STB data are provided by FourthWall Media (FWM), a commercial TV data vendor. FourthWall contracts with cable Multiple System Operators (MSOs) to install its software on cable boxes. The software records every time an event - either a change of channel, or a power off - occurs. Each device has a persistent, unique identifier such that tuning events can be (anonymously) linked to an individual device. Devices are associated with households, which are also given a unique, anonymized identifier.¹ Figure 1 presents a visualization of the tuning data for a single, randomly selected device on a single day

¹As Table 1 shows, the average household in the sample has 1.67 devices.

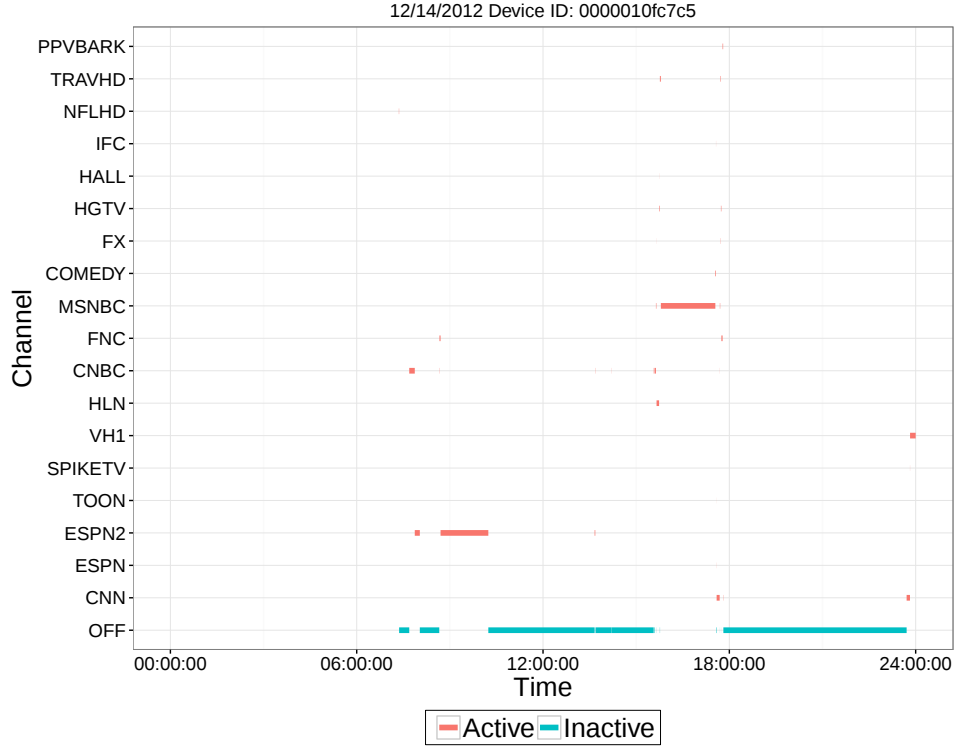


Figure 1: Visualization of STB data for a sample device on a single day.

(12/14/2012).

The data cover the time period from May 2012 to January 2013. Table 1 shows summary statistics for the data set.² The devices included in the data set are not a sample: they are the population of set-top boxes for the MSOs with which FWM has contracts. As a result, the number of devices tracked is large, with about 678K devices from just over 400K households.

	Count
Devices	677578
Households	404470
Zip Codes	996
DMAs	61
MSOs	9

Table 1: Summary statistics for FWM data, as of 11/6/2012.

²We present counts as of election day, 11/6/2012. There is some change early (May-June) in the sample period as FWM was rolling out its product during that time. Household and device counts are largely stable, and similar to November, from July 2012 on.

One limitation of the data is that FWM’s MSO partners are generally smaller systems: none of the large national conglomerates like Comcast, Cox or Charter are represented in the data. As a result, the sample skews towards smaller metro areas. The top three DMAs by device count in the data are Charleston-Huntington, WV; Bend, OR; and Wilkes Barre-Scranton-Hazleton, PA. For this reason, we also collect nationally representative aggregate data from Nielsen (described in the next section) in order to validate the temporal patterns observed in the STB data and re-weight the sample when national representativeness is required.

FWM also contracts with an external vendor to match households in its data to demographic attributes. Table 2 provides summary statistics of several key demographic variables in the sample. We use these demographic variables, along with partisanship indicators (`Democrat` and `Republican`)³ to construct an estimated Republican voting propensity in the 2008 presidential election as a baseline measure of party preference.⁴

var	min	q25	median	mean	q75	max
White	0.00	1.00	1.00	0.86	1.00	1.00
Black	0.00	0.00	0.00	0.07	0.00	1.00
Hispanic	0.00	0.00	0.00	0.06	0.00	1.00
College Grad	0.00	0.00	0.00	0.38	1.00	1.00
Age	18.00	44.00	54.45	54.14	64.00	99.00
Democrat	0.00	0.00	0.00	0.24	0.00	1.00
Republican	0.00	0.00	0.00	0.10	0.00	1.00
R Vote Propensity	0.00	0.32	0.59	0.53	0.72	1.00

Table 2: Summary statistics for FWM demographic data.

On average, the sample skews right relative to the national population. As can be seen in Figure 2, however, there is a second mode of Democratic-leaning households in the sample,

³FWM’s data vendor provides two partisanship variables, one from the head of household’s voter registration and one from self-reports. We use the value from the voter registration file wherever possible.

⁴We construct this estimate for each household in the data using a combination of the zip-code level aggregate Republican presidential vote share, and individual demographics and party affiliation. We used data from the 2008 Cooperative Congressional Elections Study (CCES) to fit a model of vote choice on demographics and party affiliation plus zip code fixed effects; the estimated coefficients from this regression were then used, along with zip code average vote shares, to predict vote probabilities for each household in the sample.

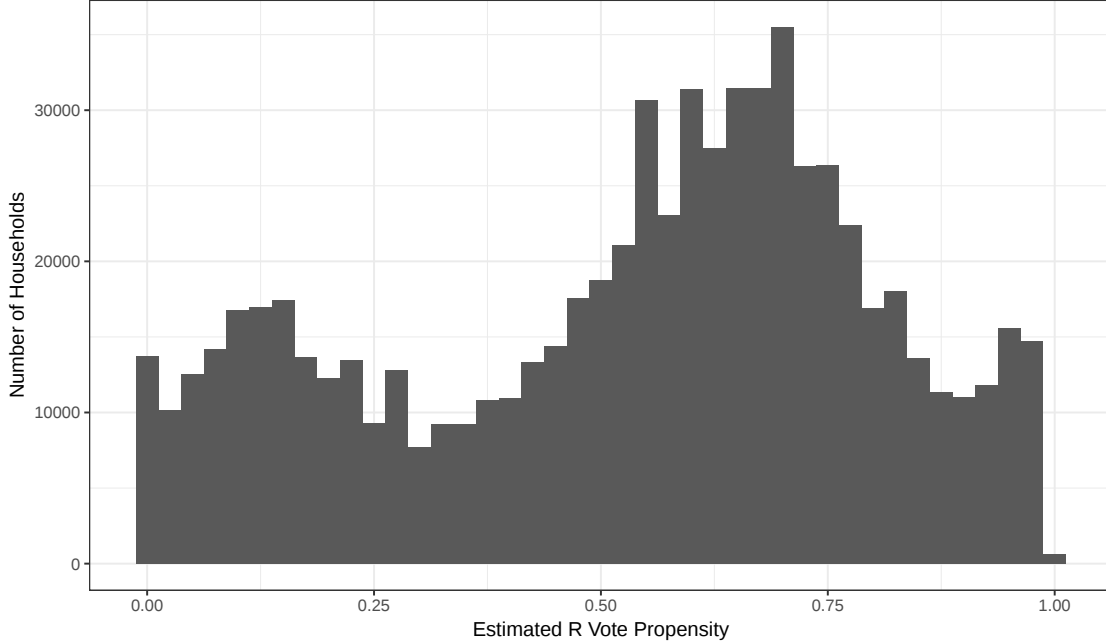


Figure 2: Histogram of estimated household-level Republican voting propensity.

and the distribution covers the entire 0-1 range.

From the raw tuning event data, we construct a dataset of ratings measured in fifteen-minute intervals for the 5PM-11PM time window, for a total of 24 fifteen minute blocks on each day in the sample period. We measure ratings at the channel-time block level as the fraction of households in the sample with devices that were active (meaning able to record tuning events) on that day who watched the channel for at least 5 minutes in the block.

Figure 3(a) plots the time series of daily primetime ratings for the three cable news channels in the FWM data. There is a substantial rise in ratings for all three channels in the two months leading up to the election, with a large spike on election day (the highest-rated day in the sample period for all three channels).

In addition to the over-time variation, we can also use the viewership data to construct differences in audience composition across shows. Figure 4 shows, for each show, the average, 25th percentile, and 75th percentile estimated Republican voting propensity of viewers of the show. All FNC shows have audiences that are, on average, more Republican than all MSNBC shows, with all CNN and network shows lying somewhere in the middle. However, there is

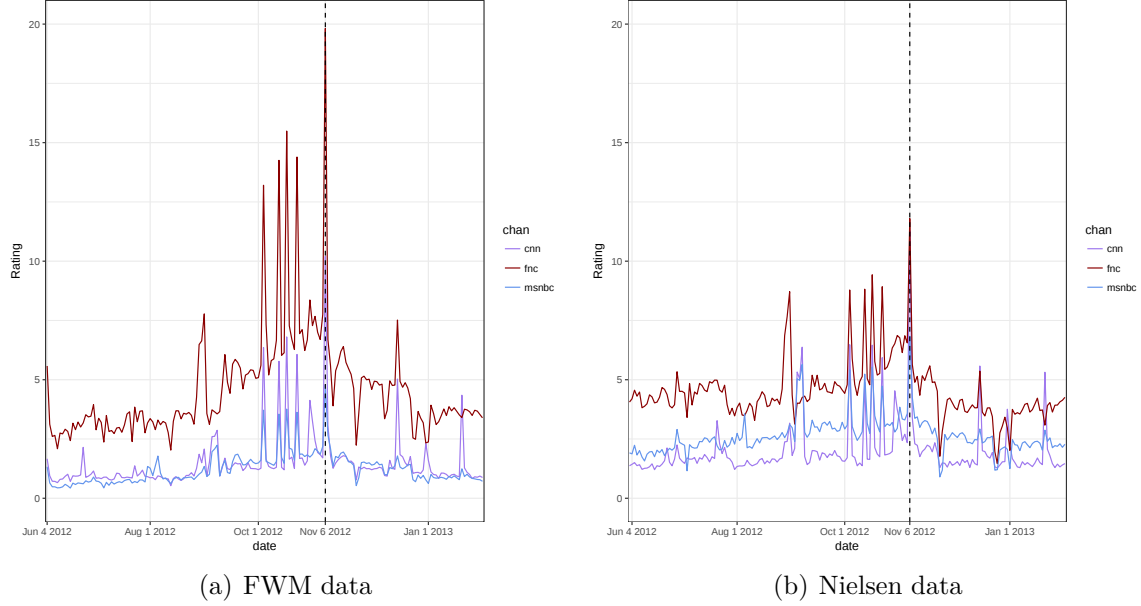


Figure 3: Ratings of news channels over time, June 2012 - January 2013. The dashed line is election day.

substantial audience overlap, with the 75th percentile viewer of every MSNBC show falling to the right of the mean viewer of every FNC show. While there is some within-channel variation in audience composition, most of the differences are at the channel level. The difference between the most Democratic MSNBC show (*Politics Nation with Al Sharpton*) and the most Republican FNC show (*The Five*) is about 17 percentage points, which is about equal to one standard deviation in the data. Hence, consistent with work such as Gentzkow and Shapiro (2011), we find nonzero but limited ideological segregation in media consumption; even the most right-wing FNC show has an audience that is only about 8 percentage points more likely to vote Republican than the average household in the data.

Finally, we examine the interaction of over-time and cross-sectional variation by exploring differential dynamic patterns by partisanship in the data. Figure 5 shows smoothed viewership of the three cable news channels over time, broken into three partisan categories: Democrats, Independents and Republicans. All three groups display a pattern of increased interest in news close to the election, with viewership rising across all three channels. The

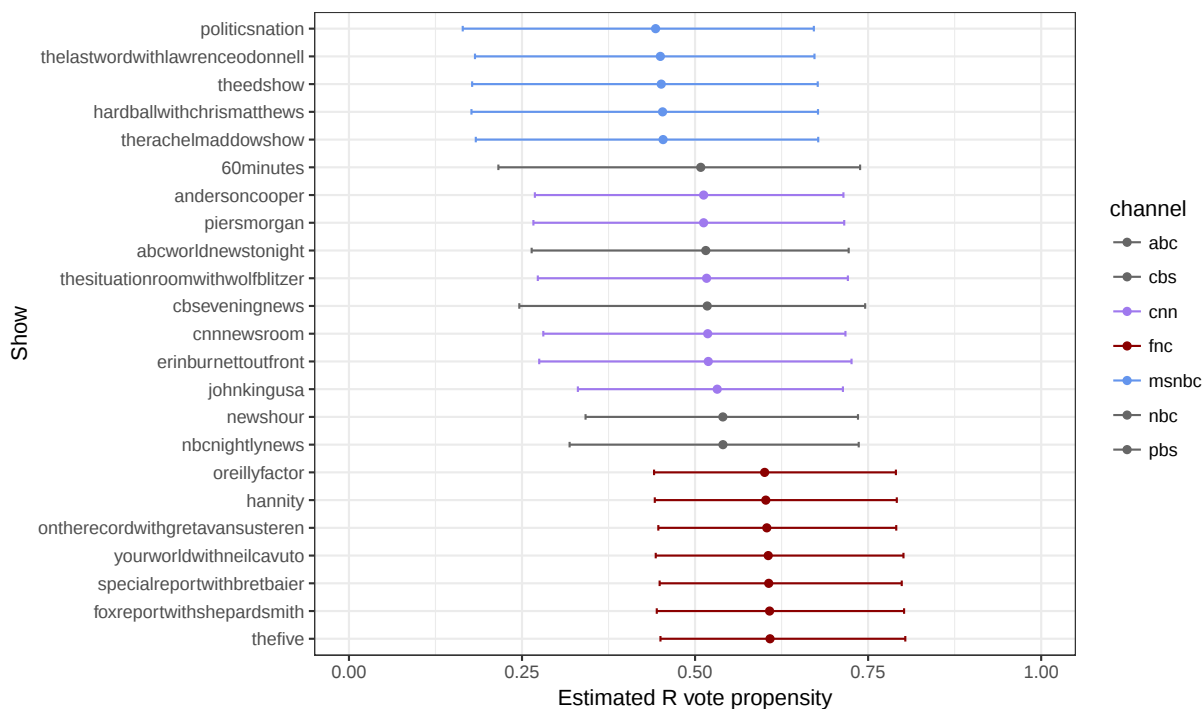


Figure 4: The average estimated Republican presidential voting propensity of viewers of each show in the data. Viewers are defined as those watching at least 5 minutes of the show on a given day. Daily statistics are averaged, weighting by the number of viewers on that day, to construct the show-level statistics plotted above. The dots show the mean estimated Republican voting propensity of viewers of each show; capped lines give the 25th to 75th percentile range.

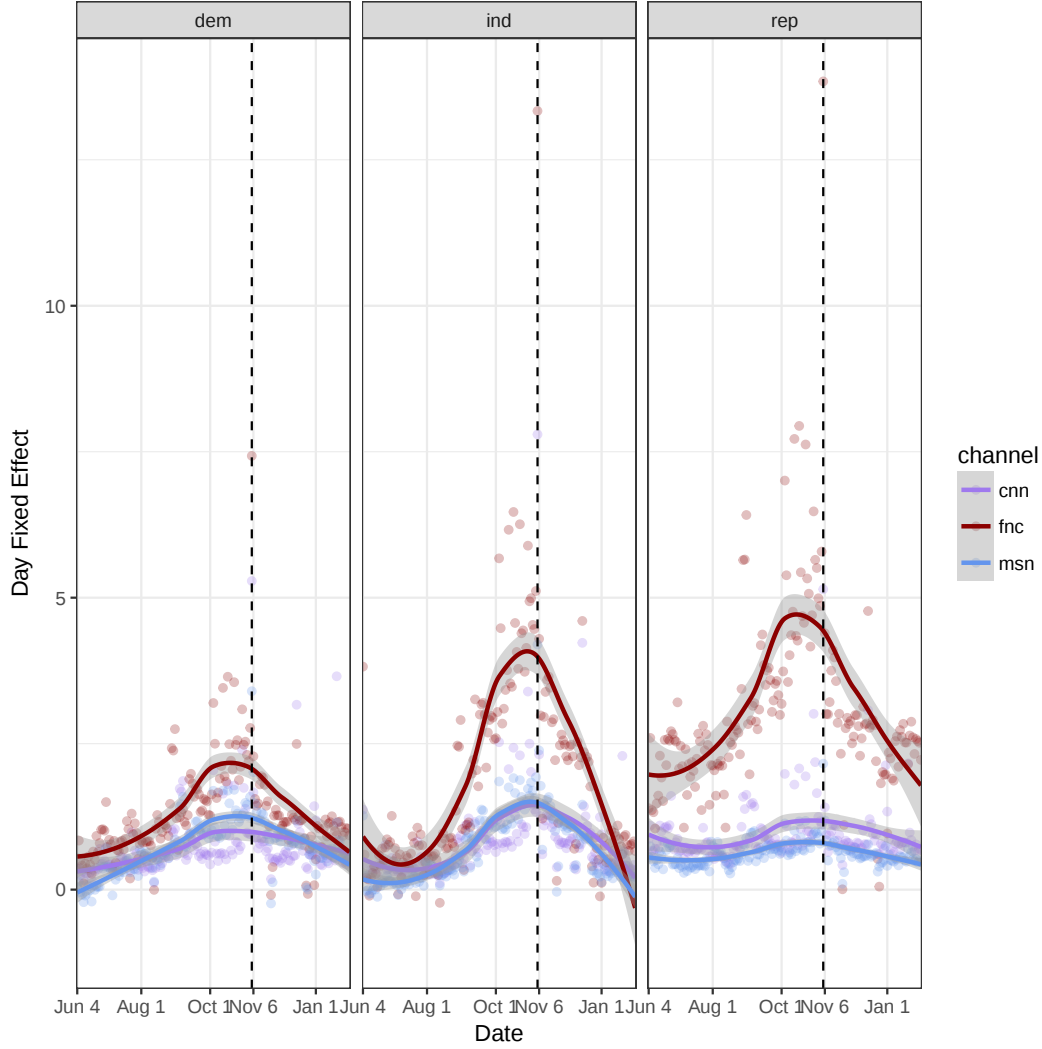


Figure 5: Daily viewership of the three cable news channels in the set top box data, by party. The lines are locally-weighted regression smoothers.

slope of the election-related surge, however, is highest for the Independents relative to the established partisans. This pattern is consistent with the theoretical perspective that instrumental information value is highest for voters for whom information acquisition might plausibly change their vote.

Aggregate Ratings Data We acquired aggregate ratings at the level of the Designated Market Area (DMA) from the Nielsen Company. The Nielsen ratings data covers the three cable channels for the time period from March 2012 to January 2013 in fifteen-minute incre-

ments. Nielsen ratings data are available for the largest 56 DMAs. Nielsen uses automated data collection hardware installed in select, randomly sampled households to gather ratings.⁵ There are a median of 1439 households per DMA in the Nielsen aggregates, with a range from 2,882 in the largest market (Los Angeles) to 431 in the smallest (Fort Meyers-Naples, FL).

Figure 3(b) plots the time series of ratings for the three channels, using Nielsen data. The time pattern is quite similar to that observed in the FWM data. Nielsen-measured ratings appear to be slightly higher for CNN compared to the FWM ratings, and substantially higher for MSNBC.

Cable Transcripts We downloaded transcripts for all shows available in the Lexis-Nexis database of news transcripts in 2012-2013. Transcripts cover all early evening and primetime weekday shows on the three cable news networks, plus the national evening news broadcasts on ABC, CBS, NBC and PBS.⁶ The transcripts indicate the time the episode aired, the speaker at each point in the program, and a transcription of what was said. We extract both the identity of speakers and the content; speakers are classified as either the show host, another network contributor (such as a field reporter), or a guest. Among guests, we create sub-categories for guests who are elected officials from either of the major parties.

We split the transcripts for each show into 15 minute blocks using a method described in Appendix A. In every time block of transcript data, we construct the frequency of every two-word phrase, or bigram. We select a subset of bigrams which are sufficiently common and appear on sufficiently many shows, and input these to a latent Dirichlet allocation (LDA) topic model using the method of Hoffman et al. (2010). Details of this procedure are in Appendix A.

⁵We use Nielsen’s “Local People Meter (LPM)” ratings, which are only available in the top 56 DMA’s; in smaller markets Nielsen collects ratings using either an older automated system (called “Set Meter”) or manual diaries recorded during sweeps weeks, neither of which have the requisite fine time resolution available for comparison with the STB data.

⁶These programs are *ABC World News Tonight*, the *CBS Evening News*, *NBC Nightly News*, and *PBS NewsHour*. With the exception of PBS’ broadcast, which airs from 6-7PM Eastern, these programs are generally a half-hour in length and begin at 7:30PM Eastern.

Label	Average Weight	Most Indicative Phrases
Filler	0.24	“web site”, “peopl know”, “want talk”, “know know”, “know peopl”, “video clip”, “good see”, “realli want”
ACA	0.10	“republican parti”, “tea parti”, “health insur”, “right wing”, “care act”, “afford care”, “obama care”, “insur compani”
Pres. Campaign	0.08	“mitt romney”, “governor romney”, “romney campaign”, “obama campaign”, “bain capit”, “swing state”, “47 percent”, “battleground state”
Debt Ceiling	0.06	“debt ceil”, “john boehner”, “hous republican”, “harri reid”, “govern shutdown”, “capitol hill”, “speaker boehner”, “immigr reform”
Sandy	0.05	“report tonight”, “hurrican sandi”, “new jersey”, “still ahead”, “staten island”, “expert say”, “news world”

Table 3: Topics with highest average weight, 2012-2013

LDA assumes the existence of a latent set of topics in a collection of documents (here, TV news segments) and produces an estimate of the probability of usage of each bigram by each topic. These probabilities, in turn, can be used to predict the probability of each document in the collection having come from each topic. We use these document topic probabilities as our measure of segment content.

The topic model produces topics that are generally quite coherent and easily identifiable. Table 3 shows the top (stemmed) phrases for each of the topics which have the highest average weight in the data.⁷ The first topic, which in addition to the phrases listed contains a number of generic connective phrases, we label as “filler” and use as the excluded category in our regression models with topic weights as predictor variables. The second most common topic clearly refers to the Affordable Care Act (ACA) and the tea-party-led efforts to repeal and replace the law. The third topic is presidential general election campaign coverage - an additional topic, not listed in the table because its average weight is lower, captures coverage of the primary campaign.⁸

The distribution of topic weights is far from uniform; the top five topics in Table 3, as

⁷We compute topic weights for each segment, then average these across all segments in the 2012-2013 period.

⁸The full set of topics, with indicative words for each, is presented in Appendix table A1.

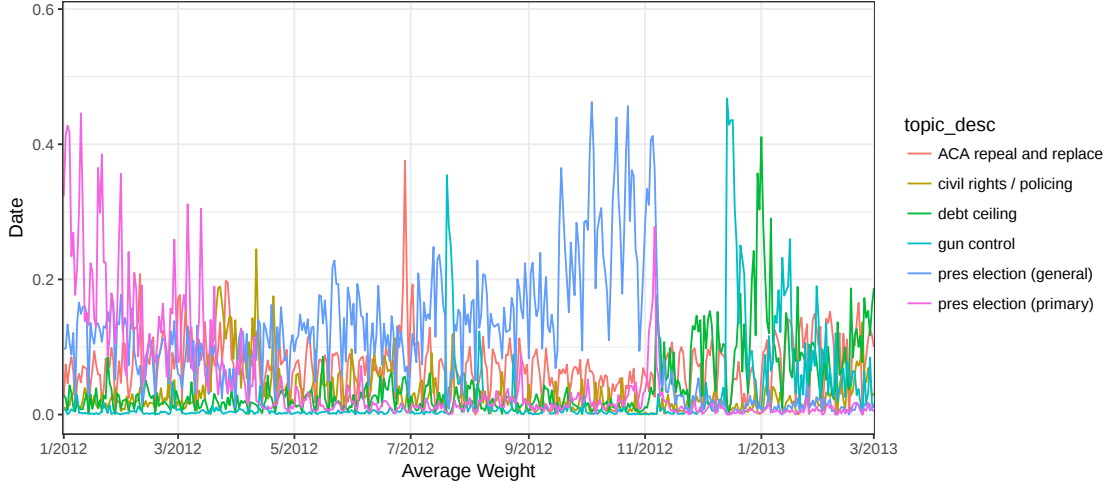


Figure 6: Time series of topic weights, for five of the top ten non-filler topics in the data.

can be seen in the second column of the table, account for more than 50% of total weight. The top 11 topics account for 75% of total weight, and the top 17 account for 90%. The bottom 28 topics account for only 5% of weight in the data.⁹

Of course, the sample averages miss substantial over-time variation in topical coverage. Figure 6 shows the time series of a few major topics’ average coverage weights by day over the course of 2012 to early 2013. The presidential primary topic dominates in early January 2012 during the Iowa caucus and New Hampshire primary, then quickly disappears in favor of the general election topic. Discussion of the ACA and Republican repeal efforts spikes in late June 2012, then rises again after the election when it became clear this would be a focus of House Republicans’ agenda. Discussion of gun control spikes in December 2012 following the mass shooting at Sandy Hook elementary school. Discussion of civil rights and policing issues shows a peak corresponding to the Trayvon Martin shooting in spring of 2012. Unsurprisingly, topics of news interest vary over time as events arise.

Coverage is, of course, also not uniform across channels. We use the coverage weights to extract two dimensions for each topic, a vertical “importance” dimension and a horizontal

⁹There are four topics which essentially never have positive weight in the transcripts during the time period for which we have viewership data; these four topics all consist of miscellaneous “grab bag” phrases and have no clear interpretation. We exclude these four topics from our models.

“slant” dimension. We construct these dimensions by regressing each topic’s coverage weights for a given channel and time-block on date fixed effects plus channel dummies. The average value of the date fixed effect for a given topic measures the persistence of its importance in news coverage across all channels.¹⁰ The slope of the channel coefficients for each topic with respect to the partisanship of the outlet¹¹ measures the topic “slant:” a topic that appears equally on all channels has slant 0, whereas a topic that appears more often on FNC would have positive slant and a topic that appears more on MSNBC would have negative slant.

Figure 7 shows the results of this exercise. The horizontal dimension shows the estimated “slant” of each topic we recover from the transcripts; topics are arranged vertically in increasing order of slant. Bolded topics are in the top quartile of “importance;” these topics feature heavily in coverage across all channels throughout the period.

Some topics are unsurprising: the topic labeled “alleged Obama scandals” contains, among other things, discussion of the Obama IRS’ supposed targeting of conservative groups for audits. This topic appears much more often on FNC than on MSNBC. The “foreign policy” topic here is heavy on discussion of Benghazi, and hence also leans strongly rightwards. Others are less so: MSNBC devotes substantially more time to presidential campaign coverage than does FNC, perhaps reflecting the campaign environment in 2012 where Obama was the incumbent and led in the polls throughout the campaign.

Finally, some topics tend to occur together in coverage; for instance, discussion of the Trayvon Martin shooting case may lead naturally to broader discussions of racial bias in policing, gun control, and so on. Table 1 shows five pairs of topics that frequently occur together in the same segment. The rightmost column is the correlation between the weight on the first topic and the weight on the second topic in the data.

¹⁰This procedure scores topics highly which have significant coverage on many days in the sample, compared to topics (such as coverage of Hurricane Sandy) that are very heavily weighted in coverage for a few days but disappear in the rest of the time period.

¹¹For purposes of constructing this measure, we set ideology of all FNC shows to 1, all MSNBC shows to -1, and all other shows to 0. Given Figure 4, this simple method is a close approximation to using the “Republican-ness” of the audience of each show to compute the slope.

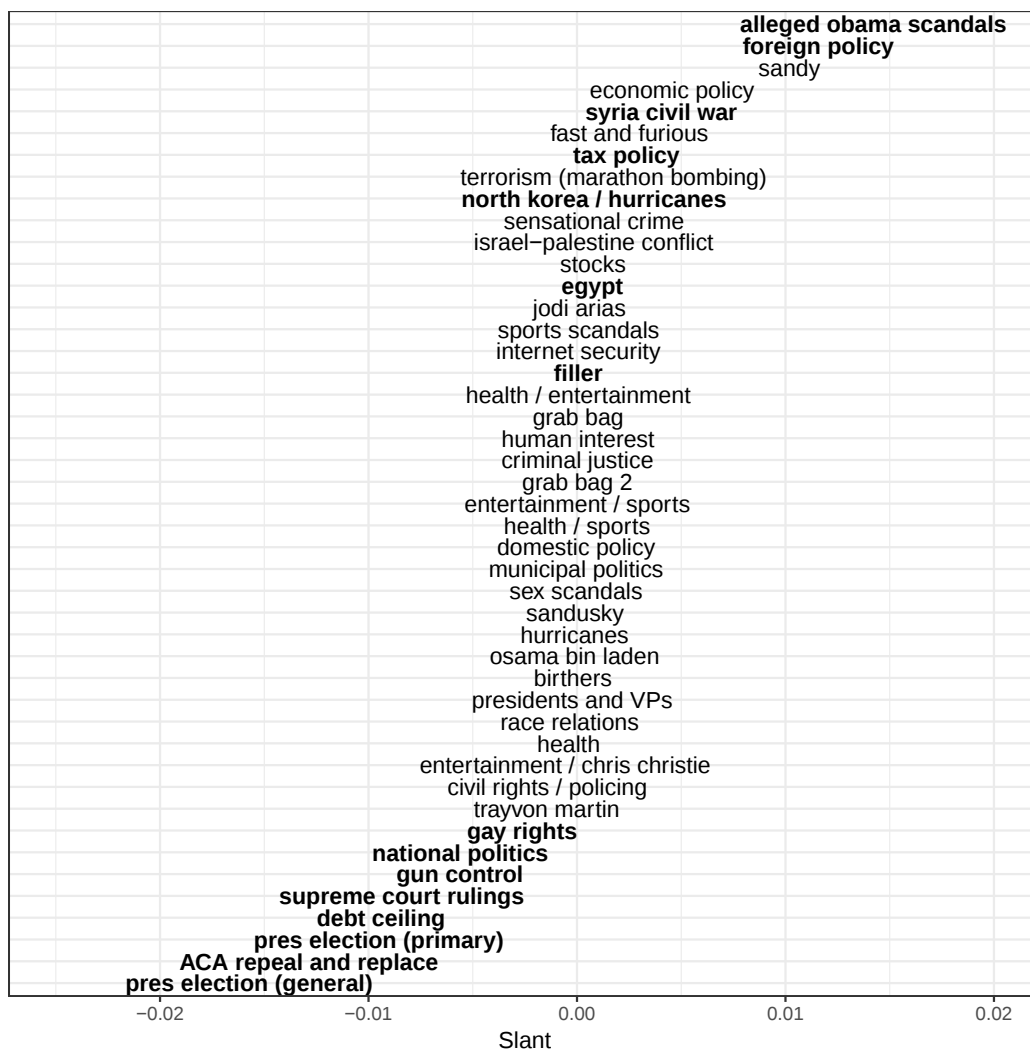


Figure 7: Topics ranked from left to right by “slant,” the slope of coverage with respect to the show audience’s political orientation. Bolded topics are in the top quartile of “importance,” the average coverage weight across all channels once day fixed effects are removed.

Topic 1	Topic 2	Corr.
civil rights / policing	trayvon martin	0.45
debt ceiling	tax policy	0.25
national politics	pres election (general)	0.19
pres election (primary)	pres election (general)	0.17
supreme court rulings	gay rights	0.15

Table 4: Topic pairs with high correlation.

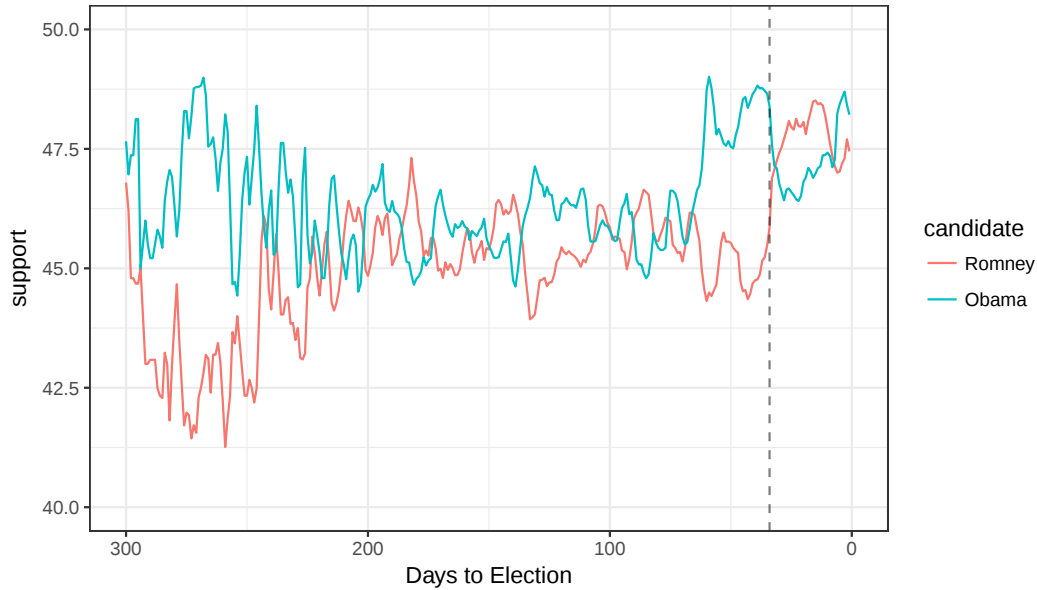


Figure 8: Polling averages over the course of the presidential campaign. The dashed line is the date of the first Obama-Romney debate on October 3, 2012, widely seen as a Romney victory.

Polling We downloaded national and state level presidential polling averages from the Huffington Post’s Pollster database. Polling data covers the 2012 election period. Pollster aggregate polls from multiple sources, weighting by sample size, to construct a daily moving average.

Polling data is useful for inferring the political implications of news stories. Figure 8 shows the national average polling support for each of the two major party candidates over the course of the campaign. The dashed line indicates October 3, 2012, the date of the first Obama-Romney presidential debate, which was widely perceived as a Romney victory. The large jump in Romney support immediately following the debate (which also corresponded with a large spike in cable news ratings visible in Figure 3) implies that the information revealed through coverage of the debate was favorable to Romney.

Figure 9 overlays coverage of Mitt Romney’s infamous “47%” comments to a group of donors:

There are 47 percent of the people who will vote for the president no matter what... who are dependent upon government, who believe that they are victims. ...These are

people who pay no income tax. ...and so my job is not to worry about those people. I'll never convince them that they should take personal responsibility and care for their lives.

All three cable channels picked up the story when it broke in mid-September, but there is a clear difference in coverage: MSNBC continued to cover the story for most of the month, whereas FNC and to a lesser extent CNN coverage quickly dropped off after the initial event. CNN and MSNBC both covered the story in significantly greater volume than FNC, which, given the audience differences between the channels, suggests negative implications for Romney. Polling data is consistent with this implication as well, as Romney's poll standing fell following the initial coverage and did not recover until the first presidential debate.

2 Model

In this section, we present a model of viewer demand for political information over the course of a campaign. The fundamental elements of the model are an unknown political state, which determines voter preferences over candidates, and a collection of news topics which may be more or less highly correlated with the political state. Viewers want to learn the political state but also have direct tastes for the “newsworthiness” or “surprise” of a news story, which is a function of the degree to which the report differs from the viewer's prior but not of its informativeness about the political state. In each time period, news arises related to each of the topics, outlets select which news items to cover, and viewers decide what, if anything, to watch. After watching, they update beliefs over each topic, and the process repeats in the next period.

2.1 Primitives

The model contains two types of actors: households, and TV channels. There are N households, indexed by i , and C channels, indexed by c . Households and channels each observe (noisy) signals about a set of $\mathcal{T}$ underlying states indexed by τ . States evolve according to

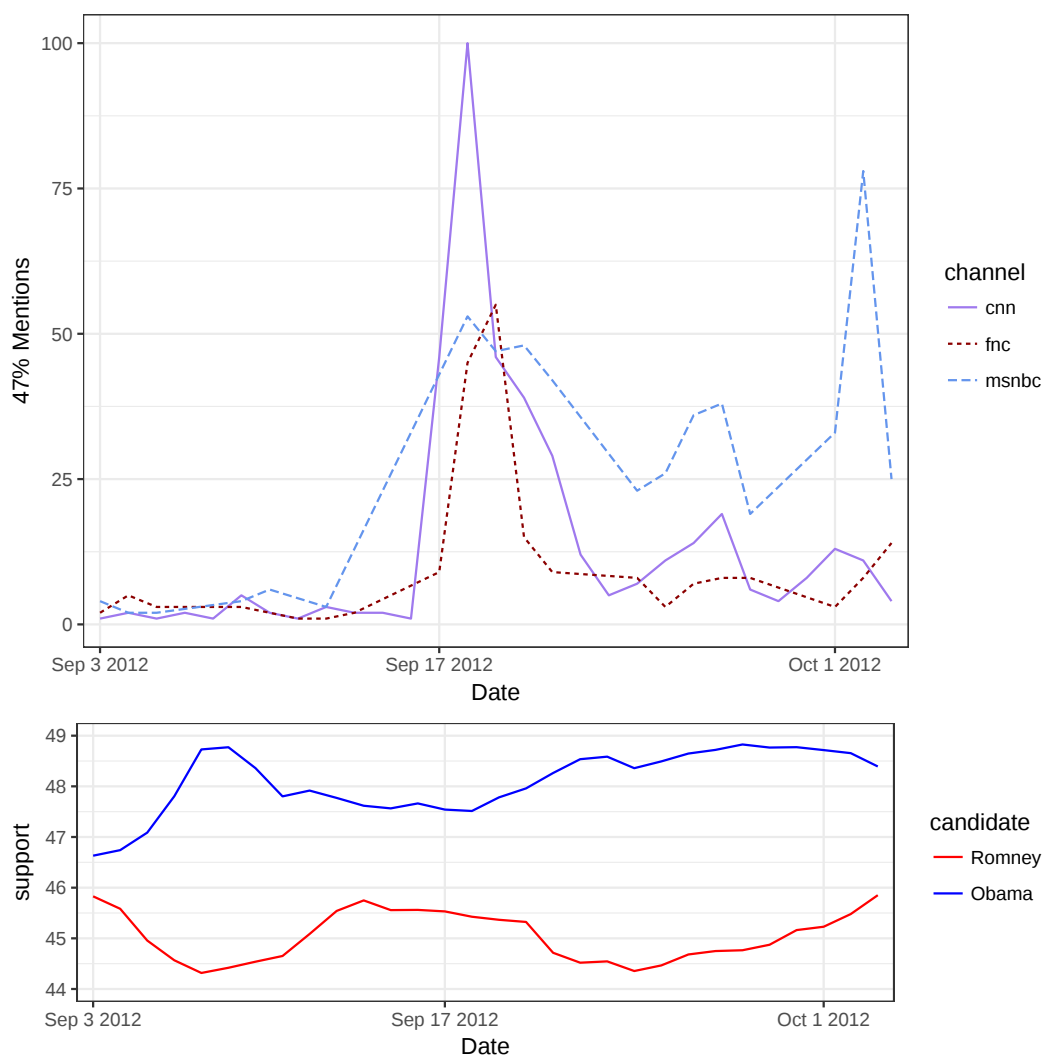


Figure 9: Coverage of Romney “47%” comment, and Romney poll standing, Sep-Oct 2012.

an exogenous random process; channels choose what to cover and viewers choose what to watch. At some fixed date, viewers make a voting decision that depends on the value of the state at that date. We describe each of these elements in turn.

Topics and States Our model of campaign information consists of a set of $\mathcal{T}$ unknown states $\{\omega^\tau : \tau \in 1, 2, \dots, \mathcal{T}\} \in \mathbb{R}^\mathcal{T}$, which may evolve over the course of a campaign. We refer to the τ 's as “topics;” ω^τ is the state variable corresponding to topic τ . Each topic-specific state ω^τ evolves each day t according to a random walk with no drift:

$$\begin{aligned}\omega_t^\tau &= \omega_{t-1}^\tau + \xi_t^\tau \\ \xi_t^\tau &\sim N(0, \sigma_\tau^2) \\ E[\omega_t^\tau] &= \omega_0^\tau \equiv \mu_\tau\end{aligned}$$

The μ_τ 's are topic “slants:” they give the a priori expected value of a report on the topic. Topics may be right-biased ($\mu_\tau > 0$), left-biased ($\mu_\tau < 0$) or neutral ($\mu_\tau = 0$).

The σ_τ^2 's give a topic's over-time variability. σ_τ^2 determines the persistence of information: how useful is knowledge of the state of topic τ today for predicting its state at some future date?

$$V(\omega_{t+k}^\tau | \omega_t^\tau) = k\sigma_\tau^2$$

These topic-specific states together form a composite political state $\bar{\omega}$, which will determine voter preferences over candidates in the election. $\bar{\omega}$ is defined as a weighted sum of the topic-specific states:

$$\bar{\omega}_t = \sum_{\tau=1}^{\mathcal{T}} \lambda_\tau \omega_t^\tau \tag{1}$$

The λ_τ 's are topic “importance” weights, which determine the usefulness of information

on the topic for estimating the current composite state. Higher λ_τ 's indicate that the state of a topic is more important for determining voting preferences, and vice versa. Weights are normalized such that $\sum_\tau \lambda_\tau = 1$ and $\lambda_\tau \geq 0$ for all τ .

Finally, every topic has some direct utility or “leisure” value $l_{i,\tau}$, which can vary by household. The population-average leisure value is $\bar{l}_\tau$. Leisure utility is described in the following section on viewer preferences.

Each topic is thus characterized by the 4-tuple $(\lambda_\tau, \mu_\tau, \sigma_\tau^2, \bar{l}_\tau)$.

Households Households make viewing decisions in every period, and a voting decision at one period denoted by t_E . While the channels can accurately observe the topic states in each period, households observe only a noisy “free” signal $\tilde{\omega}_{i,t}^\tau$ of the topic state that day. To learn more, viewers must watch the channels’ reports.

$$\begin{aligned}\tilde{\omega}_{i,t}^\tau &= \omega_t^\tau + \zeta_{i,t} \\ \zeta_{i,t} &\sim N(0, \sigma_\zeta^2)\end{aligned}$$

Voting happens once, at time t_E . There are two candidates, L or R , over which each viewer makes a binary voting decision $\nu_i \in \{L, R\}$. Viewers are characterized by an ideological position x_i , which defines utility from voting:

$$u_i^V = \mathbb{I}(\bar{\omega}_{t_E} + x_i \geq 0) \mathbb{I}(\nu_i = R) + \mathbb{I}(\bar{\omega}_{t_E} + x_i < 0) \mathbb{I}(\nu_i = L) \quad (2)$$

Voting in the model is “expressive” in the sense that the outcome voters care about is their vote choice rather than the elected candidate. Voters differ in their assessments of what circumstances would make a vote for candidate R preferable to a vote for candidate L . The larger is x_i , the more overwhelming must the evidence be that L is the correct choice, and vice versa. Some voters may hold extreme enough ideological positions ($x_i \rightarrow \pm\infty$) that they

effectively prefer one or the other candidate regardless of the state. Nonetheless, all agree that rightward movements in the composite state $\bar{\omega}$ weakly favor voting for R and leftward movements weakly favor voting for L .

Each household is endowed with a prior belief over the true value of the persistent states at time 0, namely:

$$w_0^\tau \sim N(\mu_\tau, \sigma_{\tau,0}^2) \quad (3)$$

Which implies that initial uncertainty over the value of the state at time t_E is $\sigma_0^2 + t_E \sigma_\tau^2$.

¹²

In addition to the one-time voting decision, households make every day a set of viewing decisions. We split the day into B time-blocks, indexed by b . Viewers choose one, and only one, of the C available channels¹³ or the outside option of not watching news in each time block.

Viewers get a leisure utility of watching a report on some topic determined by the topic's leisure value and its "newsworthiness" today, i.e. the size of today's innovation to the topic state relative to their prior.

$$u_{i,t,\tau}^L = l_{i,\tau}(\tilde{\omega}_{i,t}^\tau - \hat{\omega}_{i,t-1}^\tau)^2 \quad (4)$$

Where $\hat{\omega}_{i,t-1}^\tau$ is i 's estimate of the topic state ω^τ prior to viewing, and $\tilde{\omega}_{i,t}^\tau$ is i 's "free" signal today. The expected surprise for some topic is thus increasing in the difference between the prior mean and the free signal realization for that topic today.

A third component of viewing utility is from the slant of the topic, which is a function of the squared distance of the report to the viewer's political ideology:

$$u_{i,t,\tau}^S = \lambda_\tau E_{\xi_t^\tau}(\hat{\omega}_t^\tau - x_i)^2 \quad (5)$$

¹²The choice of initial mean implies that viewers' beliefs are initially correct in expectation.

¹³When we take the model to the data, actual choices are not binary, as viewers may watch more than one channel in a given time block. We treat the fraction of time spent watching each channel in a block as the empirical analogue of the model's predicted choice probabilities.

Finally, viewers face a constant “switching cost” u^W which they pay if the channel they choose at time block b differs from the channel they chose at time block $b - 1$. This cost captures the “stickiness” of viewing observed in the data.

We model voters as making a series of static viewership decisions. Voters are forward looking only in the sense that the information utility depends on how many days away the election is. Voters do not account for the fact that they will likely gather more information about the topics as the election nears. We make this modeling simplification because the computational burden of fully dynamic voters, in the context of the rest of the model, is prohibitive.

Channels and Reporting Each day, news organizations observe the vector of states ω_t (equivalently, the vector of shocks ξ_t^ω). In every time block b that day, they choose a report $\mathbf{r}_{c,t,b}$, which is a vector of length $\mathcal{T}$, $r_{\tau,c,t,b} \in [0, 1]$ for all τ and $\sum_\tau r_{\tau,c,t,b} = 1$. Channels’ objective is to maximize viewership, given constant costs of reporting on each topic:

$$\max_{\mathbf{r}_{c,t,b}} V(\mathbf{r}_{c,t,b}, \mathbf{r}_{c',t,b}) - \sum_\tau r_{c,t,b,\tau} \gamma_{\tau,c} \quad (6)$$

The model is thus one of “hard” information. Channels cannot fabricate reports, but they can select which topics to report and which to suppress each period. Viewers do not observe the channel’s report exactly: the signal they see is

$$\tilde{\omega}_{c,t,b}^\tau = \omega_t^\tau + \frac{1}{r_{\tau,c,t,b}} \epsilon_{c,t,b} \quad (7)$$

Where $\epsilon_{c,t,b}$ is an exogenous normal shock with variance σ_R^2 . The reporting weights r scale this variance, such that viewers of channel c at time block t, b observe the state with greater precision for topics on which channel c places high weight in block t, b .

Viewers combine all four components of viewing utility when computing viewing decisions. The topic-specific components of the leisure utility are weighted by the reporting weights to compute an aggregate leisure utility $u_{i,t,c}^L$. Voters treat signals that they observe

as iid draws from a normal distribution centered at the location of the topic state at the time of viewing and with variance given by $\sigma_R^2/r_{\tau,c,t,b}$. The updating process for each topic is thus a standard Kalman filter.¹⁴ Updating on the aggregate state simply involves weighting the topic-specific updated posterior means by λ_τ and the topic-specific updated posterior variances by λ_τ^2 .

2.2 Implications

Given the random walk process generating the state realizations, a viewer's optimal voting decision at time t_E involves voting on the basis of her current estimate of the composite state, e.g., voting R if $\hat{\omega} + x_i > 0$ and L otherwise. The information value of viewing a report on topic τ is then just the amount by which the additional information provided by the report changes expected utility from voting at time t_E . This depends on the reduction in uncertainty over the location of the composite state at time t_E :

$$V(\omega_{t_E}^\tau | \tilde{\omega}_t^\tau) = (t_E - t)\sigma_\tau^2 + \frac{1}{r_{\tau,c,t}^2}\sigma_R^2 \quad (8)$$

$$V(\bar{\omega}_{t_E} | \tilde{\omega}_t^\tau) = \sum_{\tau'} \lambda_{\tau'}^2 V(\omega_{t_E}^{\tau'} | \tilde{\omega}_t^\tau) \quad (9)$$

$$\Delta V = - \sum_{\tau} \lambda_\tau^2 \frac{V_t(\omega_{t_E}^\tau)^2}{V_t(\omega_{t_E}^\tau) + V(\omega_{t_E}^\tau | \tilde{\omega}_t^\tau)} \quad (10)$$

Where $V_t(\omega_{t_E}^\tau)$ is the variance of the viewer's prior at time t , including all signals received up to that point. E.g., $V_0(\omega_{t_E}^\tau) = \sigma_0^2 + t_E\sigma_\tau^2$. Finally, the change in expected voting utility derived from watching is:

$$\Delta u_{i,t}^V = \text{abs}(F_1(x_i) - F_0(x_i)) \quad (11)$$

¹⁴This is an approximation, as a fully Bayesian viewer with an understanding of the channel's objectives could update on all topics, not just the topics reported on, after viewing a report. E.g., seeing that a channel is covering, for example, Fourth of July parades likely implies that nothing more important is currently happening. The computational complexity generated by this more sophisticated form of updating is, unfortunately, infeasible for our large-scale empirical application.

Where F_1 is the CDF of a normal with mean $\hat{\omega}_{i,t}$ and variance $V(\bar{\omega}_{t_E}|\tilde{\omega}_t)$ and F_0 is the CDF of a normal with mean equal to the previous estimate $\hat{\omega}_{i,t-1}$ and the prior variance $V_t(\bar{\omega}_{t_E})$. For purposes of computing $\hat{\omega}_{i,t}$, which depends on the actual value of the report, we assume voters substitute their free signal for the actual report. If they actually watch the channel, they update based on the actual report $\tilde{\omega}_{c,t,b}^\tau$.¹⁵

Equation 11 implies two comparative statics on viewer behavior. First, the instrumental value of information increases as time approaches t_E from the left, as the closer is the date to t_E , the fewer days there are for the state to change and thus the larger will be the amount of updating generated by viewing a signal today.

The top panel of Figure 10 demonstrates this comparative static in data simulated from the model. We simulate a simple case where there are two topics, one with $\lambda_1 = 1$ and the other with $\lambda_2 = 0$, and two channels, one of which always covers topic 1 and the other of which always covers topic 2. We assume the average leisure taste for topic 2 is much higher than that for topic 1; we think of this as a simplified case of a choice between politics (topic 1) and entertainment (topic 2) news on two specialized outlets. Figure 10 shows the average expected voting utility improvement from watching the politics channel, $\Delta u_{i,t}^V$, as time goes on. The instrumental utility ramps up as the election (at time $T = 100$ in this example) approaches.¹⁶

The bottom panel of Figure 10 shows viewers' resulting channel choices in the same simulation. Initially, most voters prefer the entertainment channel (shown in orange) to the politics channel (shown in blue); as the election approaches preferences flip, until the very last days of the election when a sufficiently large fraction of viewers have watched enough to be confident in their voting decision. Day-to-day jumps in viewing channel 1 (channel 2) correspond to days with large innovations in the topic process for topic 1 (topic 2).

¹⁵We assume viewers know the topic weights $r_{c,t,b}$ on each channel prior to viewing and use these to compute viewing utility, but see the actual report only if they actually watch.

¹⁶Utility drops off in the last day or two before the election as most viewers have watched politics coverage recently and are sufficiently informed to not need to keep watching. The length and steepness of this drop-off in the final days varies with the variance parameters.

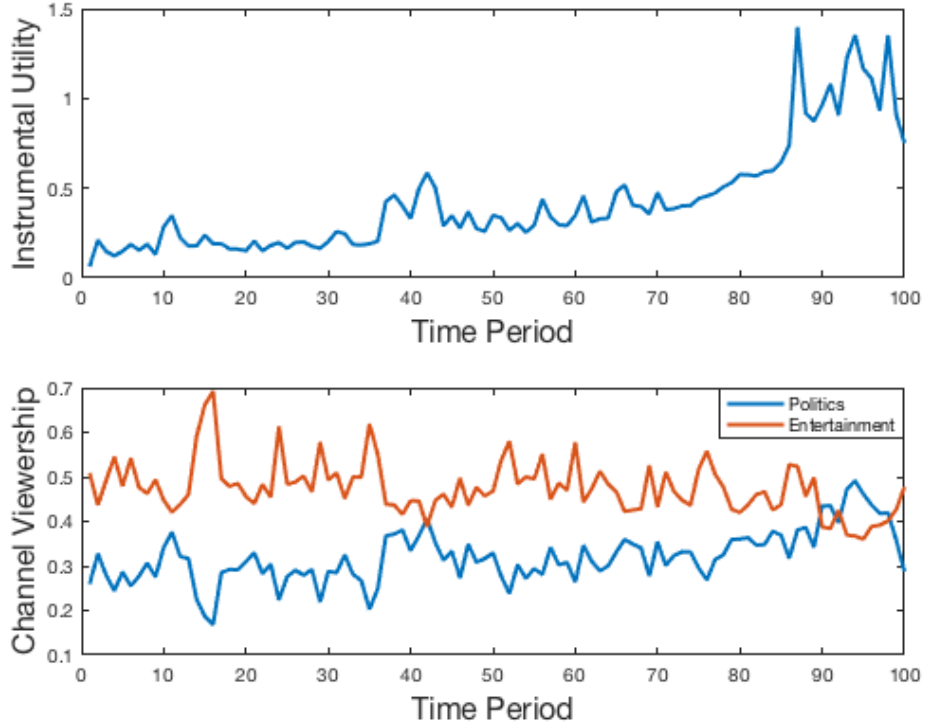


Figure 10: Top panel: simulated instrumental voting utility of watching news coverage over the course of a 100-day period. The election occurs at $T = 100$. The blue line is the expected voting utility improvement from watching, $\Delta u_{i,t}^V$, averaged across 3000 simulated viewers. Bottom panel: simulated viewing choices over a 100-day period. The election occurs at $T = 100$. The blue line is average viewership of the “politics” channel, and the orange line is viewership of the “entertainment” channel.

Second, extremist viewers (those whose x_i lies in the region where the normal CDF F is nearly flat) have little to gain from learning about the state. The likelihood that new information will cause such viewers to change their vote is very small, and hence the vote decision will be of nearly equal expected quality whether or not they watch a report. Figure 11 shows this pattern in the simulated data. We plot the expected voting utility improvement from watching the politics channel at time $t = 95$ against the voter’s initial Republican voting propensity.¹⁷ Note that the parameters and simulated topic path in this simulation lead to a landslide Democratic victory; the voter who is indifferent between the candidates at $t = 95$ has initial R voting propensity of about 0.7. The instrumental utility of watching the politics channel is highest for voters who are closest to this indifference point, and declines monotonically as we move further away from that point in either direction.¹⁸

3 Regression Results

Before moving to estimation of the full model, we present some regression estimates of relationships between quantities in the data.

3.1 Viewership

We first examine the relationship between topic coverage and channel viewership in the nationally representative (Nielsen) ratings data. We regress channel ratings (measured as the fraction of households watching the channel) on the channel’s topical coverage weights plus fixed effects for date and show dummies. The model takes the form:

$$r_{cdt} = \alpha_{s(c,t)} + \xi_d + \sum_{\tau} w_{cdt\tau} \beta_{c\tau} + \epsilon_{cdt} \quad (12)$$

¹⁷Given a set of topic parameters, one can choose x_i that would produce any given voting propensity at $t = 0$ by inverting the CDF of the appropriate normal distribution.

¹⁸The existence of variation across voters with the same x_i arises because of differences in viewing histories; some voters have more precise priors than others and hence stand to gain less from additional viewing.

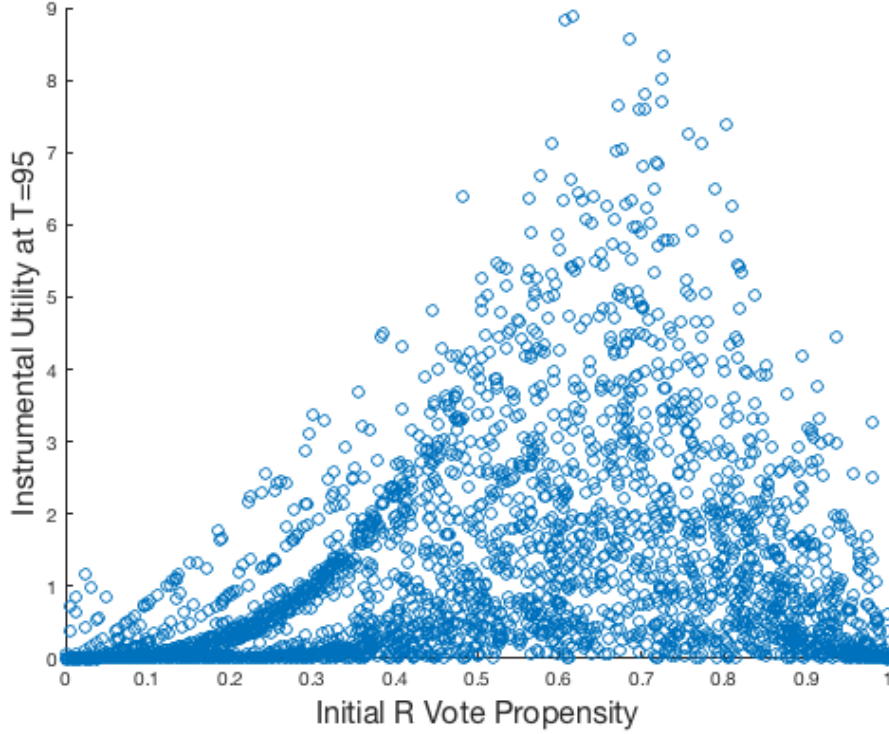


Figure 11: Expected voting utility improvement from watching the politics channel at time $T = 95$ ($\Delta u_{i,95}^V$), against initial Republican voting propensity for 3000 simulated viewers.

Where c indexes channels, d indexes days, t indexes 15-minute time blocks, and τ indexes topics. $s(c, t)$ is the show airing at time block t on show c ; $\alpha_{s(c,t)}$ is an indicator for show s , and $w_{cdt\tau}$ is the weight of topic τ on channel c , day d and time-block t . The day fixed effects ξ_d capture shocks to viewership generated by specific events, and hence the β_τ 's can be interpreted as long-term or persistent tastes for coverage of subject τ by channel c .

These persistent topic preferences differ across the channels, but generally appear to favor entertainment news, sports and crime stories. Table 5 shows the topics with the highest viewership coefficient estimates for the three cable news channels. MSNBC and CNN viewers in particular appear to like crime reporting, while FNC viewers lean somewhat more towards political stories.¹⁹

¹⁹The “Presidents and VPs” topic includes discussion of presidential and vice-presidential debates, in particular discussion making historical comparisons to past candidates. The top phrases for this topic are “vice presid,” “sarah palin,” “joe biden,” “john mccain,” “presid biden,” “dick cheney,” and “georg h.”

rank	topic	estimate	topic	estimate	topic	estimate
	CNN		FNC		MSNBC	
1	criminal justice	0.817	presidents and VPs	0.823	criminal justice	1.036
2	health	0.288	internet security	0.726	human interest	0.931
3	internet security	0.283	hurricanes	0.437	jodi arias	0.524
4	race relations	0.214	sports scandals	0.401	entertainment / sports	0.401
5	civil rights / policing	0.044	entertainment / sports	0.379	sports scandals	0.343

Table 5: Most preferred topics by channel.

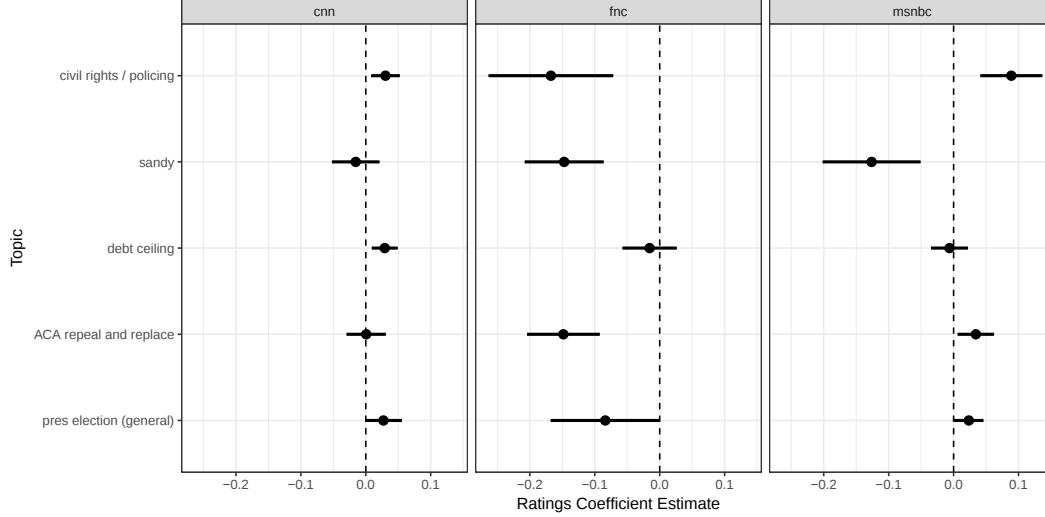


Figure 12: Viewership coefficients for five important topics on the three cable news channels.

Figure 12 shows the estimated effects and confidence intervals for a selection of the most common topics. Again, preferences differ across channels:

To capture topics which are in higher demand as the election approaches, we add interaction terms with a variable indicting the closeness to the election. The model takes the form

$$r_{c dt} = \alpha_{s(c,t)} + \xi_d + \sum_{\tau} w_{c dt \tau} \beta_{\tau} + \sum_{\tau} (w_{c dt \tau} * e_d) \beta_{\tau}^e + \epsilon_{c dt}$$

$$e_d = \begin{cases} d & d < d_E \\ 0 & d \geq d_E \end{cases}$$

Where d_E is the index of the date at which the election occurs. Figure 13 displays the

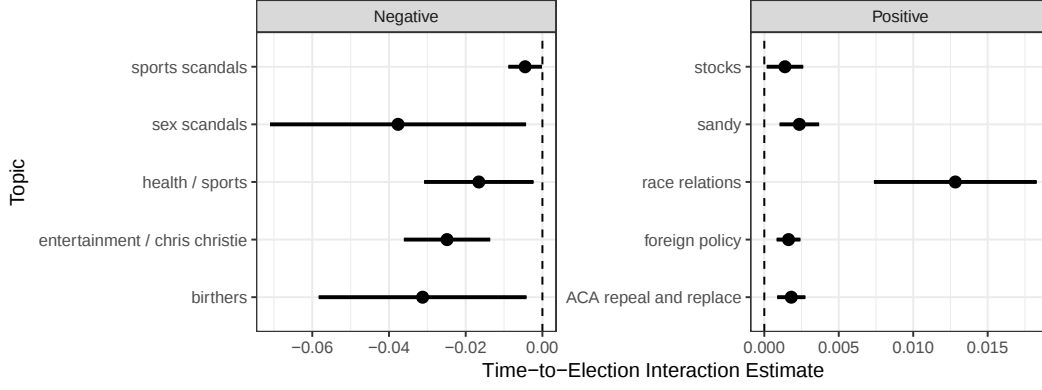


Figure 13: Coefficients and confidence intervals on interactions of topic weights with days-to-election variable. The plotted topics are the five largest and five smallest interaction terms, among the set of interactions significantly different from zero at the 5% level.

results graphically, showing estimates and confidence intervals for the topics with the most positive and most negative interaction terms.²⁰ Topics related to sports and entertainment populate the negative interaction panel, implying that these topics have abnormally low ratings in days close to the election date; the positive panel is generally made up of “hard news” topics such as the Affordable Care Act or foreign policy.²¹

We next move to estimating relationships in the household-level STB data. We estimate a series of models of the form

$$r_{icdt} = \alpha_{s(t)} + \xi_d + \sum_{\tau} w_{cdt\tau} \beta_{c\tau} + X_i \gamma_c + \delta_c r_{icdt-1} + \epsilon_{icdt} \quad (13)$$

The primary differences from the aggregate regression are the household-level viewing observations and the addition of household-level attributes X_i and lagged viewing r_{icdt-1} . Table 6 shows the estimated coefficients of δ and γ in models for each channel. X_i here consists of three variables: the estimated Republican voting propensity x_i described above, and two transformations of x_i , $(\max\{x_i - 0.5, 0\})^2$ and $(\max\{0.5 - x_i, 0\})^2$. The latter two variables

²⁰We include in this ranking only topics with interaction coefficients statistically different from zero, of which there are 15.

²¹The exception is coverage of Hurricane Sandy, which made landfall on the east coast a few days prior to the 2012 election. The timing of the storm, which aligned so closely with the election, is a challenge for our use of the distance-to-election interaction as a proxy for informativeness.

add quadratic terms for distance from the center of the range of the voting propensity variable, split into positive and negative parts to allow the curvature on each side of $x = 0.5$ to differ.

term	abc	cbs	cnn	fnc	msnbc	nbc	pbs
Lag Viewing	8.523e-01 (6.969e-05)	8.315e-01 (7.157e-05)	8.260e-01 (2.179e-05)	8.683e-01 (1.938e-05)	8.550e-01 (2.413e-05)	7.566e-01 (8.457e-05)	8.433e-01 (6.168e-05)
R Vote Propensity	-6.324e-03 (4.868e-04)	5.649e-04 (4.940e-04)	-1.443e-04 (4.108e-05)	7.762e-05 (7.744e-05)	5.516e-04 (7.324e-05)	-1.292e-04 (7.925e-04)	-1.139e-03 (1.076e-04)
R Vote Squared (-)	-1.348e-02 (8.968e-04)	4.723e-03 (7.924e-04)	3.588e-04 (9.683e-05)	-7.374e-03 (1.710e-04)	5.386e-03 (1.193e-04)	-1.342e-02 (1.274e-03)	-3.523e-03 (2.076e-04)
R Vote Squared (+)	2.020e-02 (8.562e-04)	1.842e-03 (7.573e-04)	8.083e-05 (9.272e-05)	1.540e-02 (1.633e-04)	-2.641e-03 (1.140e-04)	7.421e-03 (1.223e-03)	4.231e-03 (1.981e-04)
Hour Dummies?	Y	Y	Y	Y	Y	Y	Y
Date FE?	Y	Y	Y	Y	Y	Y	Y
N	52M	50M	636M	635M	454M	48M	98M
R squared	0.76	0.74	0.7	0.77	0.74	0.66	0.66

Notes: Standard errors in parentheses. An observation is a viewer-time block. Data is all time blocks from 4-11PM eastern time, on weekdays from 6/4/2012 to 1/31/2103. The dependent variable (and lagged dependent variable) is the fraction of time in the block spent watching the channel. The R Vote Squared (+) and (-) terms are defined, respectively, as the minimum of R Vote Propensity - 0.5 and zero and the minimum of 0.5 - R Vote Propensity and zero, squared. All regressions also include measures of topic content described below.

Table 6: Regressions of viewership on topic coverage and viewer characteristics, using household-level set-top-box data.

Table 6 shows that there is high persistence in viewing: the estimated coefficients $\hat{\delta}$ range from 0.76 (NBC) to 0.87 (FNC), implying that a viewer who spent all 15 minutes in block t watching Fox News would be expected to spend about 13 minutes watching Fox News in block $t + 1$. To aid in the interpretation of the coefficients on household Republican voting propensity, Figure 14 plots the predicted deviations from mean viewing as a function of R voting for the three cable channels. FNC’s appeal accelerates for very likely Republican voters, and decelerates for very likely Democratic voters, albeit at a slightly lower rate. MSNBC is close to flat for most of the spectrum but increases its appeal (at a smaller rate than FNC) as Republican voting propensity approaches zero. CNN viewing has essentially no relationship to viewer partisanship.

Analogously to the aggregate data regressions, we also add in some specifications interactions of $w_{cdt\tau}$ with the election time trend e_t and the individual characteristics X_i . Figure 15 displays estimated regression coefficients and confidence intervals for the interaction terms of $w_{cdt\tau}$ with the predicted household Republican voting propensity.²²

²²Because the sample in these regressions is so large and standard errors correspondingly tiny, we plot

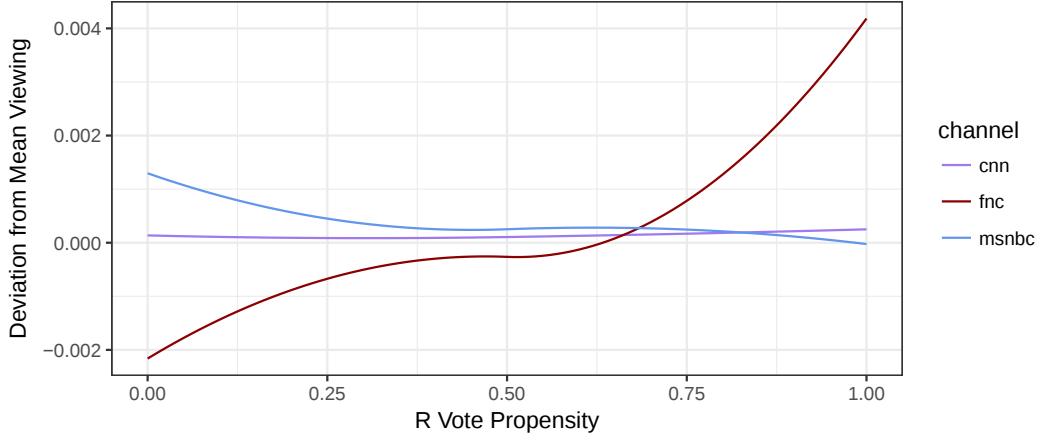


Figure 14: Predicted deviation from mean viewing time for the cable news channels, as a function of households' Republican voting propensity. Curves are plotted using the estimated coefficients from Table 6

3.2 Polling

Table 7 shows the relationship between daily polling changes and ratings on the cable channels. The dependent variable in column (1) is the change in the Obama share of respondents identifying a preference for one of the two major party candidates; in column (2) the dependent variable is the absolute value of this change. Regressors in both equations are the ratings, measured by Nielsen, of the three cable channels on the previous day. Increased viewership of FNC is correlated with worse Obama poll performance on the subsequent day. The opposite is true for MSNBC ratings, although the standard errors are much wider.

4 Model Estimation and Parameter Estimates

Estimation Procedure We estimate the model by indirect inference [Gourieroux et al. \(1993\)](#). For a given set of model parameters, we simulate a set of individuals described by their initial ideology and leisure tastes for each of the topics. We estimate analogous

confidence intervals as the range of ± 6 standard errors around the point estimate, and display the largest magnitude coefficients from among the set of interaction terms for which this confidence interval does not overlap zero.

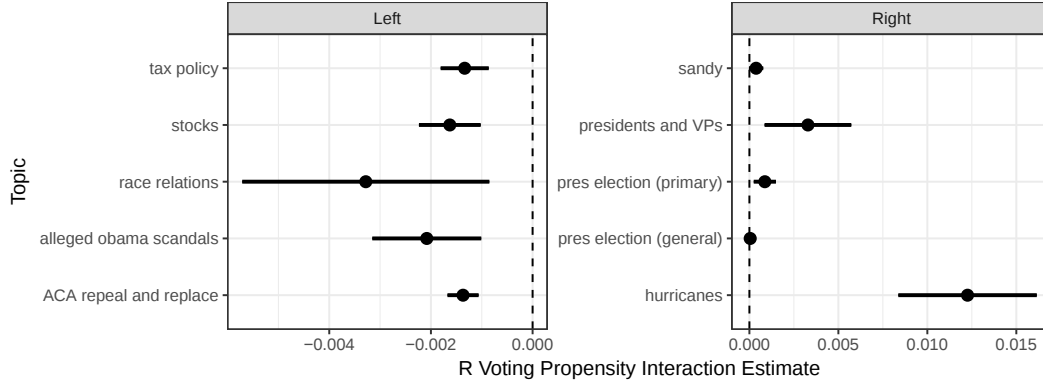


Figure 15: Interaction coefficients of topic weight with households' predicted Republican voting. We select the topic interaction coefficients with largest positive and largest negative magnitudes.

Table 7: Regressions of Polling Changes on Cable News Ratings

	Change in Obama Two Party Share (1)	Absolute Change in Obama Two Party Share (2)
Lag CNN Ratings	0.004 (0.003)	0.007 (0.002)
Lag FNC Ratings	-0.003 (0.001)	-0.002 (0.001)
Lag MSNBC Ratings	0.002 (0.003)	-0.002 (0.002)
Constant	0.001 (0.001)	0.003 (0.001)
N	109	109
R ²	0.082	0.145

Notes: Data are daily polling and average daily ratings data during the period 6/4/2012 - 11/6/2012. Dependent variable is the change in the Obama share of respondents stating a preference for Obama or Romney, in directional (column 1) or absolute (column 2) terms.

regressions of viewership and poll changes to those presented for the real data in Section 3, and choose model parameters θ to match the regression coefficients as closely as possible.

Simulation of viewership in the model of the previous section requires a realized path of topic innovations, which are unobserved. As the actual data are generated by a single realization of events (the events that actually occurred and were covered in the news from June 2012-January 2013) it would be inappropriate to integrate out the topic innovations in the model by simulating many draws and averaging viewership and polling over all draws. Instead, we estimate the most likely topic path given data on topic coverage, viewership and

polling. The estimation of the topic path occurs once per day; we choose the vector ξ_t to match time-block-level viewership and daily polling data for that day, given the choice of topic coverage by each channel in each time block.

Our estimation method alternates back and forth between estimating the day-to-day topic path given a set of global parameters, and estimating the global parameters given an estimated topic path for the entire time period. We continue this alternating procedure until convergence.

The topic path step consists of matching (1) the difference between predicted and observed polling by day, and (2) the difference between predicted and observed viewership by channel-period. Predicted viewership is computed by estimated channel choice probabilities for each simulated individual in each time block given the model parameters, the topic path, the individual’s viewing history, and channels’ reporting decisions, and averaging across viewers. Polling is constructed analogously, averaging the votes of all simulated viewers if they voted on the basis of their estimate of the aggregate state today. The optimization takes the form:

$$\min_{\xi_t} \left(\sum_{c,b} (v_{ctb} - \hat{v}_{ctb}(\theta, \xi_t, r_{ctb}))^2 + (p_t - \hat{p}(\theta, \xi_t, r_{ctb}))^2 + \sum_{\tau} \frac{\xi_{t,\tau}^2}{\sigma_{\tau}^2} \right) \quad (14)$$

Where r_c is channel c ’s observed topic coverage weights, p_t is the two-party poll share of the Republican candidate on day t , and v_{ctb} is the average viewership of channel c at time block b on day t . Hats indicate simulated quantities. The final terms in the sum are penalties for large innovations; the penalties scale inversely with the corresponding topic variance parameter, ensuring that smaller variance topics get smaller innovations on average and vice versa. Each day contains $24 \times C$ viewership moments to match, plus the polling and the penalty moments. In our application to the three cable news channels, there are thus 118 moments available to fit each day’s 45 topic innovations.

With the estimated topic path in hand, we compute a set of moment conditions for each candidate global parameter vector. The moments we match for this step consist of

regression coefficients for four regressions: (1) an individual level regression of viewership on topic weights, topic weights interacted with number of days until election, and topic weights interacted with individual Republican voting propensity, plus show fixed effects and date fixed effects (performed on a sample of the simulated individuals who match observable characteristics of the individuals in the STB data), (2) an aggregate level regression of viewership on topic weights, topic weights interacted with number of days until election, show fixed effects, and date fixed effects, (3) aggregate regressions of daily poll changes on daily topic coverage on each channel, and (4) an aggregate regression of daily poll changes on viewership. The optimization for the global parameters takes the form:

$$\min_{\theta} \sum_i w_i \left(m_i(\theta, \xi, \mathbf{r}) - m_i^0 \right)^2 \quad (15)$$

Where m_i is a regression coefficient from one of the regressions listed above, m_i^0 is the value of the corresponding regression coefficient in the actual data, and w_i is a positive weight. In our implementation, we use weight equal to the inverse of the square of the regression standard error corresponding to each regression coefficient. This weighting scheme ensures that the optimizer places more emphasis on matching moments that are precisely estimated in the data, compared to moments that are imprecisely estimated. There are a total of 755 regression coefficients (moments) in the objective function, and 231 parameters.

5 Model Output

We present here results based on a preliminary set of parameters estimated using the process described above. The optimization had not yet converged as of the time of writing, and hence these results should be treated as no more than a proof of concept. Nonetheless, even at these very preliminary estimates, the model can reasonably well match the important over-time and cross-sectional patterns in the viewership data.

First, the model captures the increasing demand for cable news as the election approaches.

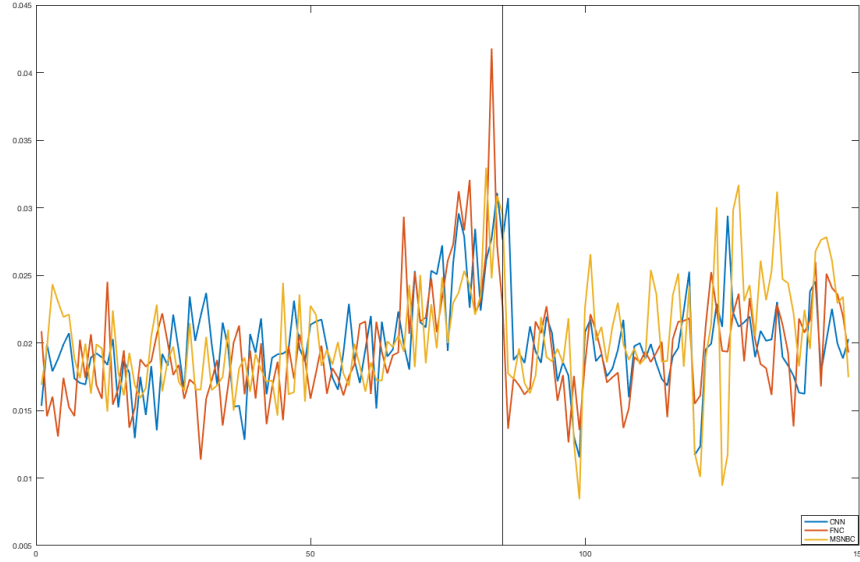


Figure 16: Time series of simulated viewing from the model. Day 1 is June 4, 2012; the vertical line indicates election day, November 6, 2012.

Figure 16 shows the time series of simulated viewership from the beginning of the period to election day. Simulated viewership begins to rise roughly a month before the election, then drops off abruptly afterwards, returning quickly to the baseline level.

Second, the model captures the differential time trends evident in the actual data: relatively centrist viewers see greater relative increases in viewing close to the election, compared to their more extreme counterparts. Figure 17 splits the simulated voters into three groups: those who initially lean Democrat (with R vote probability of less than $\frac{1}{3}$), those who initially lean Republican (with R vote probability of at least $\frac{2}{3}$) and those in the middle. We plot viewership among these three groups relative to their respective baselines (the average viewing in days 1 to 30 of the simulation). As in the actual data, centrist viewers see the largest relative surge in viewership as the election approaches.

Finally, the model generates fairly strong positive informational externalities. Viewers who have high leisure tastes $l_{i\tau}$ watch more news and, consequently, have more precise and more accurate beliefs about the location of the aggregate state on election day. In Figure

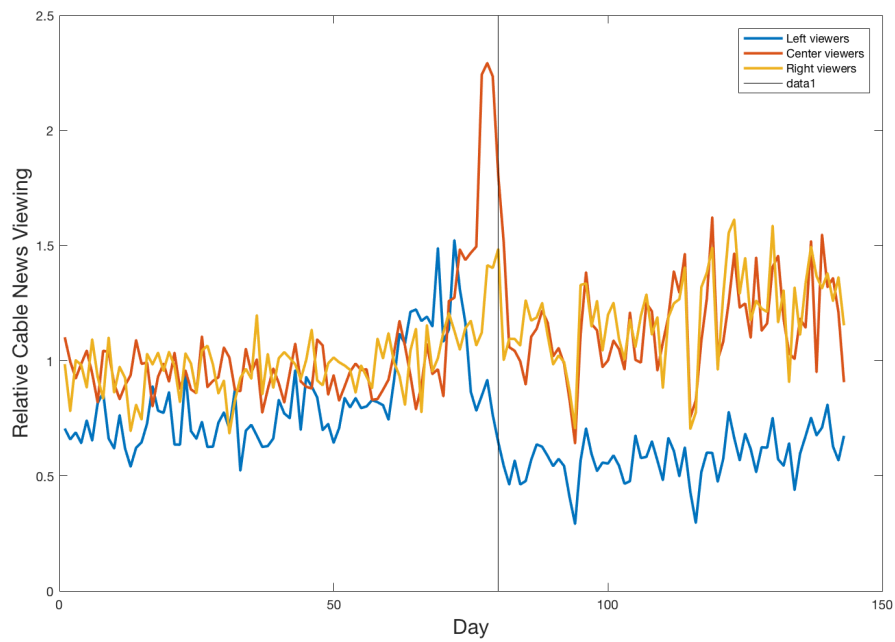


Figure 17: Simulated viewership among different voter types. “Left” viewers are those with initial R voting propensities less than 0.33; “Right” viewers are those with initial R voting propensities greater than 0.66; “Centrist” viewers are the remainder. The vertical line indicates election day, November 6, 2012.



Figure 18: The relationship between viewers’ non-instrumental “leisure” tastes for news and the variance of viewers’ posteriors about the location of the aggregate state on election day.

18 we plot the relationship between individual consumers’ leisure tastes (averaged over all topics τ) and the variance of their posterior beliefs about the location of the aggregate state on election day. The relationship is negative, with overall correlation of -0.22.

Although viewing cable news provides some information on the political state, the aggregate quality of information provided by cable news coverage is fairly poor. First, 17% of simulated viewers make the “wrong” vote choice, in the sense that they would vote differently if voting on the basis of the actual state location rather than their post-viewing estimate of the state location. Second, Figure 19 shows that channels’ informational content is generally fairly low compared to an optimally-informative benchmark. We construct this benchmark by computing, in each time period, the topic τ^* for which the product of viewers’ average posterior variance and $\lambda_{\tau^*}^2$ is largest; this is the topic which currently contributes the most to viewers’ uncertainty about the aggregate state. The benchmark reporting strategy sets

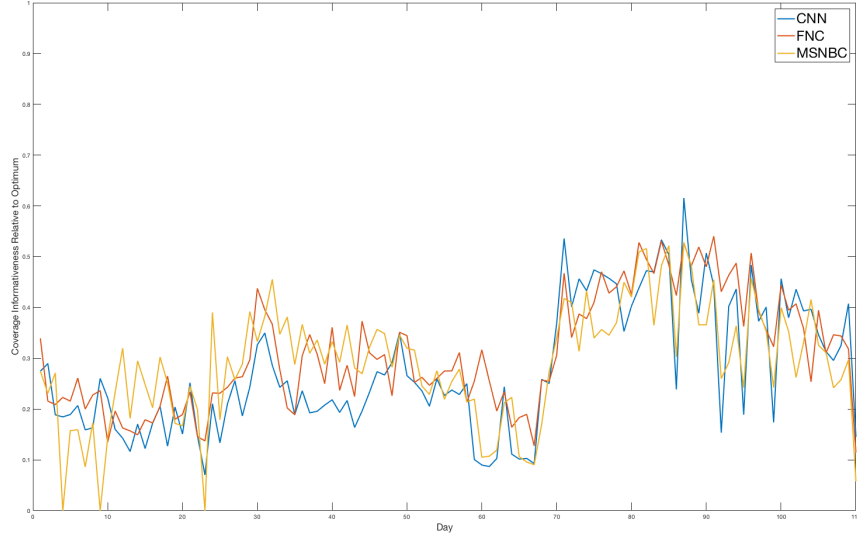


Figure 19: The informativeness of channels' coverage over time, relative to a benchmark.

r_{c,t,b,τ^*} equal to one and all other components of $r_{c,t,b}$ to zero. Channels' actual coverage rarely breaks 50% of this benchmark, and more typically hovers in the 20-30% range; informativeness actually declines in the last month prior to the election when the largest numbers of new viewers are tuning in.

6 Discussion

Demand for election news coverage on television is driven by a combination of factors: an instrumental demand for information about the candidates, taste for news as entertainment, and taste for reporting that confirms viewers' ideological priors. We systematically measure content choices in TV news programming and document the reaction of viewers to these content choices, distinguishing all three components of demand. This news-demand model can be used to understand how actual news provision provided by viewership-maximizing channels compares to optimal news provision that would be provided by a social planner concerned with the quality of voters' information at election time. Preliminary results from the model suggest that the actual provision of news coverage is far from achieving this

informational optimum.

References

- DellaVigna, Stefano and Ethan Kaplan**, “The Fox News Effect: Media Bias and Voting,” *The Quarterly Journal of Economics*, 2007, *122* (3), pp. 1187–1234.
- Durante, Ruben and Brian Knight**, “Partisan control, media bias, and viewer responses: Evidence from Berlusconi’s Italy,” *Journal of the European Economic Association*, 2012, *10* (3), 451–481.
- Ely, Jeffrey, Alexander Frankel, and Emir Kamenica**, “Suspense and Surprise,” *Journal of Political Economy*, February 2015, *123* (1), 215–260.
- Enikolopov, Ruben, Maria Petrova, and Ekaterina Zhuravskaya**, “Media and political persuasion: Evidence from Russia,” *The American Economic Review*, 2011, *101* (7), 3253–3285.
- Garcia-Jimeno, Camilo and Pinar Yildirim**, “Matching Pennies on the Campaign Trail: An Empirical Study of Senate Elections and Media Coverage,” 2017.
- Gentzkow, Matthew and Jesse M. Shapiro**, “What Drives Media Slant? Evidence From U.S. Daily Newspapers,” *Econometrica*, 2010, *78* (1), 35–71.
- **and —** , “Ideological segregation online and offline,” *The Quarterly Journal of Economics*, 2011, *126* (4), 1799–1839.
- Gottfried, Jeffrey, Michael Barthel, and Amy Mitchell**, “Trump, Clinton Voters Divided in Their Main Source for Election News,” Technical Report, Pew Research Center 2017.
- Gourieroux, Christian, Alan Monfort, and Eric Renault**, “Indirect inference,” *Journal of Applied Econometrics*, 1993, *8* (S1), S85–S118.

- Hoffman, Matthew, Francis R. Bach, and David M. Blei**, “Online learning for latent dirichlet allocation,” in “Advances in neural information processing systems” 2010, pp. 856–864.
- Kamenica, Emir and Matthew Gentzkow**, “Bayesian Persuasion,” *The American Economic Review*, 2011, *101* (6), 2590–2615.
- Martin, Gregory J. and Ali Yurukoglu**, “Bias in Cable News: Persuasion and Polarization,” *American Economic Review*, 2017, *107* (9), 2565–2599.
- Mullainathan, Sendhil and Andrei Shleifer**, “The Market for News,” *American Economic Review*, 2005, *95* (4), 1031–1053.
- Peisakhin, Leonid and Arturas Rozenas**, “Electoral Effects of Biased Media: Russian Television in Ukraine,” SSRN Scholarly Paper ID 2937366, Social Science Research Network, Rochester, NY 2017.
- Prat, Andrea**, “Media power,” *Journal of Political Economy*, 2017, *Forthcoming*.
- Strömberg, David**, “Mass media and public policy,” *European Economic Review*, 2001, *45* (4), 652–663.
- , “Mass Media Competition, Political Competition, and Public Policy,” *The Review of Economic Studies*, January 2004, *71* (1), 265–284.
- Vanderbilt**, “Television News Archive,” 2017.

A Topic Model Details

We collected text transcripts of news programs on CNN, Fox News, MSNBC, ABC, CBS, NBC, and PBS for the period 1/1/2012 to 12/31/2013. For cable news, our transcript data covers all regular programs airing between 4PM and 11PM on weekdays. For network news, our data covers the major network evening newscasts plus the *PBS Evening NewsHour* and the Sunday network shows *This Week*, *Meet the Press*, *Face the Nation* and *60 Minutes*. Transcripts were downloaded from the Lexis-Nexis news database, with the exception of the FNC show *The Fox Report with Shepard Smith*, which was unavailable in Lexis-Nexis and which we obtained instead from the Internet Archive.²³

We parsed every available transcript using a script which extracted both the name of the person speaking (a guest, an anchor, a reporter, etc.) and the transcript of their speech, split into sentences. The script then reduced all transcribed words to word stems using the Porter stemming algorithm,²⁴ removed a list of common “stop words,” and constructed frequency counts of every two-word phrase (“bigram”) appearing in the sentence.

The next step was to assign sentences to the time-block in which they would have aired. This is a non-trivial task because the transcripts do not contain time stamps, generally arriving in a single file with the full half- or one-hour-long show transcript time-stamped only with the show start time.²⁵ To assign transcripts with no segment time stamps, we used the following method.

We collected data from the Vanderbilt Television News Archive (Vanderbilt, 2017), which recorded the start and end times of every news segment and commercial block that aired on the network news programs and CNN’s 7PM news show during our sample time frame. Using this data, we constructed, for each minute of a show, the average fraction of time devoted to news (as opposed to commercial) content. The probability of a given minute containing ads

²³<https://archive.org/details/tv>

²⁴We used Dag Odenhall’s Haskell interface to the Snowball NLP library (<https://hackage.haskell.org/package/snowball>) to perform this step.

²⁵Certain FNC and ABC shows are an exception, being broken into individual segment transcripts with start times for the segment. We use this information when available.

is shown below in Figure A1; the left plot shows the pattern on CNN’s one-hour show, and the right panel shows the pattern on the networks’ half-hour shows. There is a clear, regular pattern where earlier minutes in a show are less likely to contain advertising, and hence more likely to contain news content. We used this minute-by-minute weight to construct, for each time block, the total fraction of a show’s non-commercial content that on average would have aired in that block. Those weights were then used to allocate transcript sentences to time blocks.²⁶

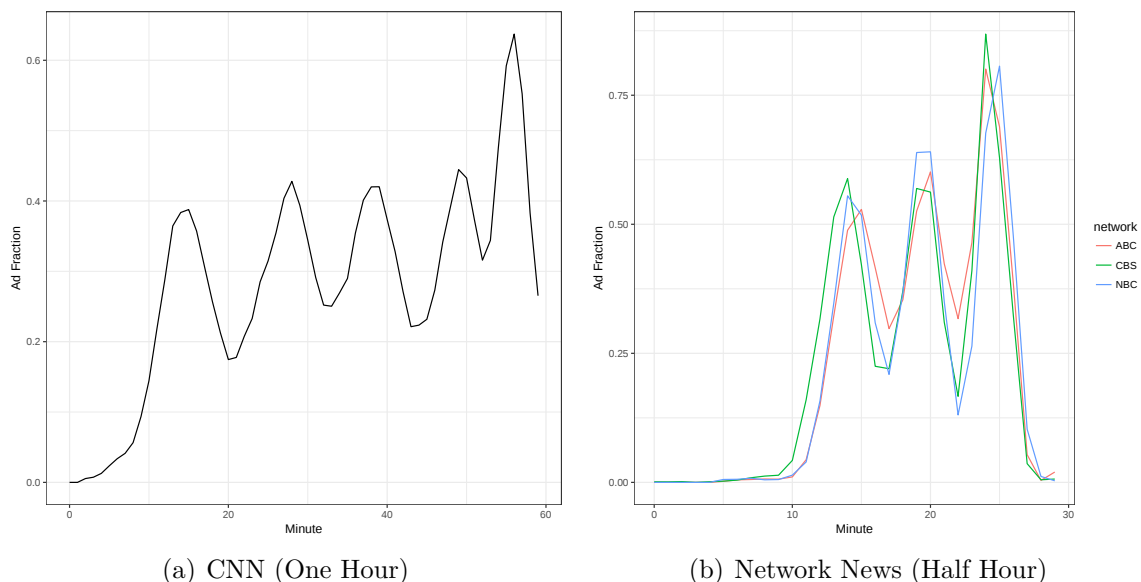


Figure A1: Advertising frequency by show minute.

Once a show’s transcript sentences were assigned to time-blocks in this fashion, we aggregated the bigram counts for each show and time-block. Because the frequency distribution features a large mass of very infrequent phrases - more than 50% of bigrams appear only once in the entire collection of transcripts - we apply some minimum frequency criteria to limit the set of bigrams input to the topic model. These criteria are that the phrase must appear:

1. At least 50 times across all shows and dates,

²⁶The exception to this rule is the PBS Evening NewsHour, which contains no ads and which we therefore allocated evenly across time-blocks.

2. And on at least 5 show episodes,
3. And on at least 3 channels,
4. But on less than 700 days.

The final criterion drops about 50 extremely common phrases such as “good night,” “good evening,” and similar. The minimum channel requirement drops many common but show-specific phrases (such as “Thing 1” and “Thing 2”, a recurring segment on *All in with Chris Hayes*). Finally, in addition to these criteria we also dropped all bigrams containing the names of one of the networks, one of the shows, or one of the anchors.²⁷

A total of 110,075 phrases survived these checks. The frequency counts for phrases in this set in all 58,086 “documents” - 15-minute chunks of transcript text - were then input to a LDA topic model which was fit using the online algorithm of Hoffman et al. (2010). We estimated a model with 50 topics, using a minibatch size of 4096 documents, 30 passes over the corpus and tuning parameter values recommended by Hoffman et al. (2010).

The full set of topics, with the top phrases for each, is given in Table A1.

²⁷So long as the name was sufficiently distinctive to ensure that the reference could only be to the show/anchor; for instance “situation room” is a common phrase referring to an important room in the White House and thus was not dropped even though it is also the name of Wolf Blitzer’s CNN show.

topic description	word0	word1	word2	word3	word4	word5	word6	word7
north korea / hurricanes	north korea	north korean	break news	kim jong	south korea	situat room	per hour	d c
israel-palestine conflict	prime minist	middl east	ceas fire	benjamin netanyahu	isra prime	minist netanyahu	minist benjamin	tel Aviv
internet security	credit card	bank account	cyber attack	first class	five week	red light	pearl harbor	secur expert
suprem court rulings	suprem court	vote right	plan parenthood	scott walker	right act	feder govern	court rule	chief justic
stocks	web site	point close	wall street	minimum wage	jeffrey brown	long term	young peopl	unemploy rate
criminal justice	physic evid	will success	cat mous	idea give	mous game	finish sentenc	power way	two name
gay rights	sex marriag	gay marriag	african american	north carolina	cathol church	gay right	marriag equal	pope franci
entertainment / chris christie	new jersey	chris christi	governor chris	whitney houston	governor christi	jersey governor	michael jackson	prescript drug
sandy	report tonight	hurrican sandi	new jersey	still ahead	staten island	expert say	news world	news new
grab bag 1	mitt romney	georg zimmerman	north korea	south carolina	newt gingrich	trayvon martin	law enforc	break news
foreign policy	state depart	secretari state	hillari clinton	al qaeda	nation secur	susan rice	foreign polici	talk point
domestic policy	food stamp	super pac	welfar reform	work requir	diseas control	center diseas	enhanc drug	perform enhanc
fast and furious	attorney general	secret servic	eric holder	justic depart	fast furious	general eric	darrel issa	depart justic
civil rights / policing	stand ground	law enforc	police depart	african american	defens attorney	ground law	grand juri	justic system
egypt	muslim brotherhood	moham morsi	middl east	presid moham	tahrir squar	prime minist	secur forc	presid morsi
filler	p m	00 p	tomorrow night	new year	10 00	town hall	9 00	00 eastern
health / entertainment	heart attack	heart diseas	abraham lincoln	gang member	civil war	peanut butter	two minut	poverti rate
presidents and VPs	vice presid	sarah palin	joe biden	john mccain	presid biden	dick cheney	georg h	h w
entertainment / sports	law school	real estat	best interest	matt lauer	angel time	tom cruiz	scienc fiction	like tom
human interest	public school	role model	audio gap	everyth will	easter egg	smell like	want grow	march band
hurricanes	new orlean	super bowl	tropic storm	gulf coast	storm surg	bobbi jindal	ann romney	hurrican katrina
health	good morn	breast cancer	morn america	georg will	georg stephanopoulos	hash tag	jinni kimmel	cancer patient
trayvon martin	georg zimmerman	trayvon martin	self defens	second degre	degre murder	sanford florida	neighborhood watch	martin famili
osama bin laden	bin laden	air forc	osama bin	sexual abus	john f	presid kennedi	kill osama	forc base
sandusky	police offic	penn state	jerri sanduski	ariel castro	joe paternno	football coach	sex abus	shot fire
sensational crime	sexual assault	death penalti	police say	harri reid	littl girl	old girl	three women	10 year
debt ceiling	john boehner	rick santorum	hous republican	harri reid	govern shutdown	capitol hill	speaker boehner	immigr reform
pres election (primary)	mitt romney	rick santorum	newt gingrich	south carolina	new hampshir	ron paul	republican parti	rick perri
terrorism (marathon bombing)	al qaeda	law enforc	9 11	boston marathon	drone strike	homeland secur	boston bomb	finish line
health / sports	sanjay gupta	medic correspond	chief medic	debit card	jeremi lin	asian american	next hour	correspond dr
municipal politics	million dollar	half million	citi council	dollar year	police forc	police chief	citi manag	peopl citi
syria civil war	chemic weapon	u n	nuclear weapon	unit nation	civil war	al assad	use chemic	red line
economic policy	wall street	gas price	john edward	student loan	billion dollar	interest rate	top stori	lot money
tax policy	tax cut	middl class	fiscal cliff	tax rate	rais tax	social secur	tax increas	spend cut
jodi arias	door open	doubl standard	door open	energ room	travi alexand	left hand	leav door	maricopa counti
race relations	south africa	nelson mandela	civil right	martin luther	luther king	african american	tip day	south african
births	donald trump	birth certif	conspiraci theori	born unit	obama born	bath salt	health depart	presid born
sex scandals	san francisco	anthoni weiner	high speed	new yorker	eliot spitzer	sex scandal	amanda knox	speed rail
ACA repeal and replace	republican parti	tea parti	health insur	right wing	care act	afford care	obama care	insur compani
alleged obama scandals	nation secur	e mail	edward snowden	phone call	jay carney	web site	tea parti	washington post
national politics	paul ryan	bill clinton	mitt romney	first lady	vice presid	presid unit	georg w	second term
sports scandals	lanc armstrong	new england	rose garden	realiti show	tour de	act terror	de franc	seven time
grab bag 2	time magazin	3 year	mother day	7 year	star trek	jennif hudson	can eat	month old
gun control	background check	gun control	assault weapon	gun violenc	sandi hook	mental health	second amend	mental ill
pres election (general)	mitt romney	governor romney	romney campaign	obama campaign	bain capit	swing state	47 percent	battleground state

Table A1: The top 8 two-word phrases associated with each topic.