The Impact of Political Campaigns on Demand for Partisan News

Luis Menéndez , Manuel Lleonart-Anguix *

November 26, 2024

Abstract

How do people acquire political information during political campaigns? Using a unique dataset that comprises both audience data and text content from Spanish TV news, we estimate the demand for political information. We rely on Large Language Models to categorize the tone associated with each political party in each story of the day. To address endogeneity concerns regarding the political leaning of the content offered, we use input shocks that constrain channels' political news production asymmetrically. While outlets strive to maintain their political stance, these shocks affect them differently, depending on the day's random news composition. Our findings show that the demand for political information decreases during political campaigns. Moreover, campaigns trigger a polarized news consumption, with right-leaning viewers demanding more favorable content on their own party and more negative coverage of the opposing parties.

1 Introduction

Media plays a critical role in the information dissemination and might have persuasive effects on voting outcomes. On the basis of this, policy makers often regulate media markets to ensure plurality of information. Direct efforts to regulate TV content during political campaigns are currently applied in several countries, often by enforcing proportional airtime rules.[1]

*Universitat Autònoma de Barcelona, CSIC and Barcelona School of Economics

[1]In Italy, the *Par Condicio law* mandates equal time and space on television, radio, and press for all political actors during campaign periods. Similarly, the *Conseil supérieur de l'audiovisuel* applies a similar rule to both private and public broadcasters in France. Spanish "Ley de Medios" recommends equal time rule policies only for the public broadcaster Televisión Española (TVE).

However, to understand the effectiveness of these policies, it is crucial to evaluate how viewers respond to the content that is being offered. In addition, previous work has showed media influence on voting behavior during political campaigns (Enikolopov et al., 2011). This is particularly relevant in the context of Spain, which is at the top of the European countries in terms of political polarization according to recent surveys (Edelman, 2023). In particular, there has been a significant rise of affective polarization during our sampling period [2]. It is then natural to wonder whether the political news consumption patterns of the citizens exhibit similar patters.

In this paper, we use a unique dataset that comprises both the supply and demand of Spanish TV news to estimate preferences for political content using a revealed preferences approach, thus overcoming some of the well-known problems associated with survey designs (Prior, 2009).

Our focus is on testing how this demand changes during political campaign periods. Previous works have highlighted that the increasing trend towards soft news has contributed to a declining interest in news (Patterson, 2000) . Our findings are in line with those in Gambaro et al. (2021): periods where politics are combined with other type of content seem to make people engage with politics.

We then examine the existence of polarized news consumption. Using different text analysis techniques, we control for the political stance and duration of the stories (Puglisi and Snyder, 2015). Specifically, we rely on Large Language Models (LLMs) and feed all our stories to ChatGPT-4, asking it to discern the different tones associated with the various political parties. The results for tone across different outlets are consistent with their viewers' preferences as reported by survey data and enable us to discern between left, right, and middle outlets in terms of the parties they support the most.

We abstract from demand or supply-driven considerations (Gentzkow et al., 2011) and assume that outlets play a purely horizontal game where they vary the content they produce in order to capture the market. We estimate a discrete choice model (Berry, 1994) where individuals choose their preferred channel as a function of the content it offers.

Under this scenario, the political tone that the outlets offer is an equilibrium outcome and thus endogenous in the demand estimation. To address this concern, we use random variations in the inputs of the channels. Specifically, we obtain and classify all the daily stories produced in Spain during our sample period from a large news provider and treat them as the set of stories available to the TV outlets. Given that channels have established their political stance in the differentiation game and it is costly to depart from them in the short run, random variation in the political composition of the day constrains left and

[2] We refer the reader to Centro de Estudios Murciano de Opinión Pública (CEMOP) (2024) for results on affective polarization.

right-leaning channels' production decisions differently.

Previous works Durante and Knight (2012) found that viewers respond to changes in media content driven by ownership shifts. Our results indicate that viewers polarize their news consumption diets during political campaign periods even though the supply of partisan content becomes more homogeneous across outlets. Although other different forms of selectivity might be present, in this work we limit to attention to partisan selectivity in TV news.

Our results are consistent with previous findings in the US (Peterson et al., 2017). In the pre-campaign period, there is no asymmetric demand for political content between right- and left-wing audiences, neither when controlling for intention to vote nor by channel-specific viewers. During political campaigns, however, we find evidence of affective polarization. More right-wing viewers demand more negative content on the opposing left party and more positive content on their own. This shift in the demand behavior occurs parallel to the outlets becoming more homogeneous in their political stances. Furthermore, there is evidence of partisan divide for the taste of political content overall.

The fact that the echo chamber formation occurs only during campaign links our findings to existing theoretical works on information acquisition. Our findings oppose a model of instrumental voters where information acquisition helps them make better voting choices (Larcinese, 2003). Under this scenario, the value of political information (and thus its demand), should be higher as elections approach.

This work contributes to the existing literature on political preferences in media content by exploiting exogenous variation in the news landscape to identify political preferences. In particular, our setup is flexible and relies on unsupervised classification techniques that make political classification straightforward. To the best of our knowledge, we are the first paper to examine preferences for political content during political campaign periods while addressing the endogeneity in content supply.

The rest of the paper is organized as follows. In section 2, we briefly summarize the Spanish political and TV landscape. Section 3 introduces our data and describes the text analysis techniques we employ, along with some descriptive statistics on both the content and audience sides. In section 5, we describe our market setup and consider the estimation of soft vs. hard news. Section 4 discusses the endogeneity problem and the relevance of our instrument. Results from the demand estimation are shown in section 6 Finally, section **??** shows the current work in progress where we try to incorporate and estimate the supply side.

2 Context

2.1 Spanish TV market

Television is still the primary source of information in most European countries (Parliament, 2024). Figure 1 shows the main media used to acquire political information for different European countries, as reported by the 2023 Eurobarometer, broken down by age groups. Television is the preferred medium for almost all age cohorts, a pattern that is even stronger in countries such as France and Italy.

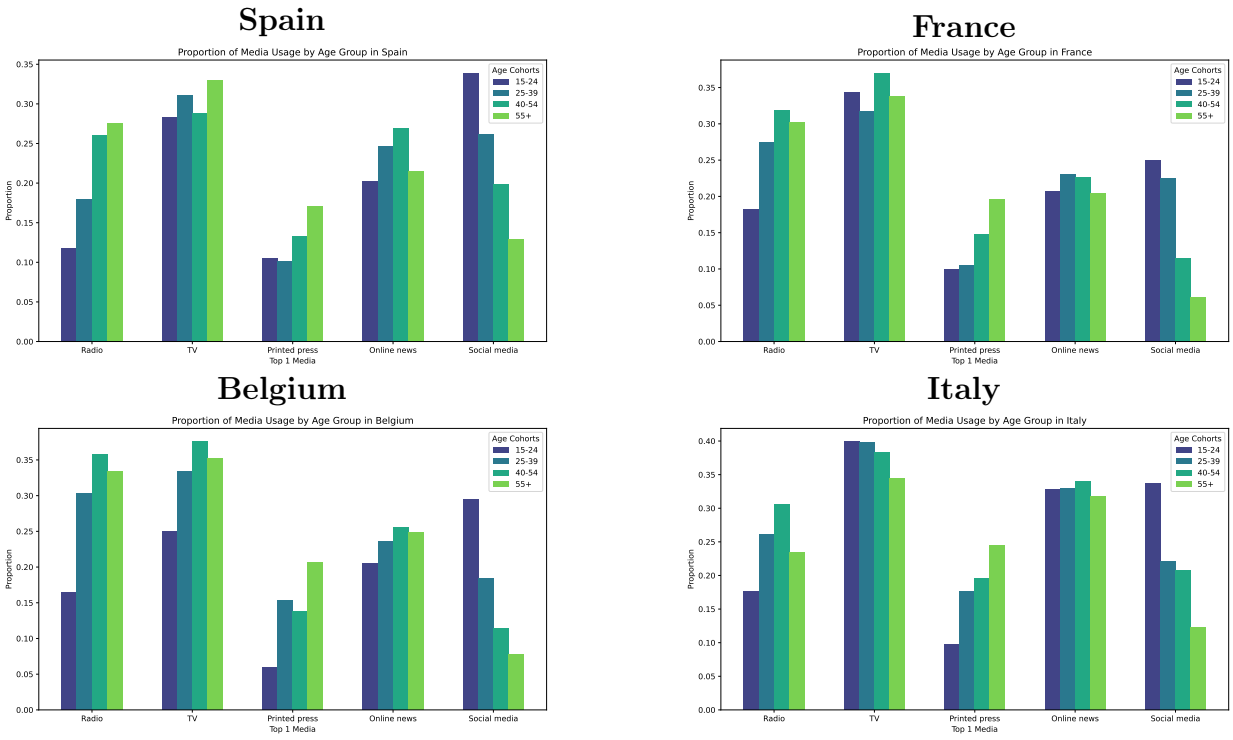

Figure 1: Histogram on the preferred media used for political information in Spain. Source: Eurobarometer, 2022.

The Spanish TV market is a competitive mix of public and private broadcasters, with Televisión Española (TVE) positioned as the state-owned provider of a variety of news, cultural programs, and entertainment. The two primary private conglomerates, Atresmedia and Mediaset, dominate most of the market. Atresmedia operates Antena 3 and LaSexta, while Mediaset controls Telecinco. The mean audience for the channels considered in this study is 2.6 million viewers per day—half a million more than the largest U.S. broadcaster alone: Fox News.

We focus only on the night edition of TV news programs for several reasons. First, it is the so-called "golden time" of the day, attracting the maximum number of viewers on TV. Second, even though channels offer other political programs, the homogeneity of TV news allows us to make very clean comparisons. These programs are broadcast every day at

almost the same time and all share a very similar structure, with a presenter introducing the main stories of the day. The well-known informational motives of these shows ensure that people seek to get informed but do so through the outlet that treats news in their preferred way, either due to some perceived outlet quality or differences in the way content is presented.

Altogether, the night editions of these TV news programs capture around 50% of the market share, equating to 8 million viewers, which represents 23% of the potential voting population. For comparison, the average number of viewers for the most popular TV news program in the U.S., Fox News, is around 1.7 million.

2.2 Political orientation of the audience

Figure 2 shows the correlation between political orientation and preferred channel for acquiring political information, based on survey data from the Centro de Investigaciones Sociológicas (CIS). The figure confirms the brief description outlined above. Right-wing individuals (PP-VOX) tend to watch Antena 3 more, whereas left-leaning individuals are divided between LaSexta and the public channel TVE. We can also observe that the correlation with Telecinco initially appears counterintuitive: it attracts viewers from both extremes of the political spectrum, supporting the hypothesis that third factors, such as entertainment, may influence their choice.

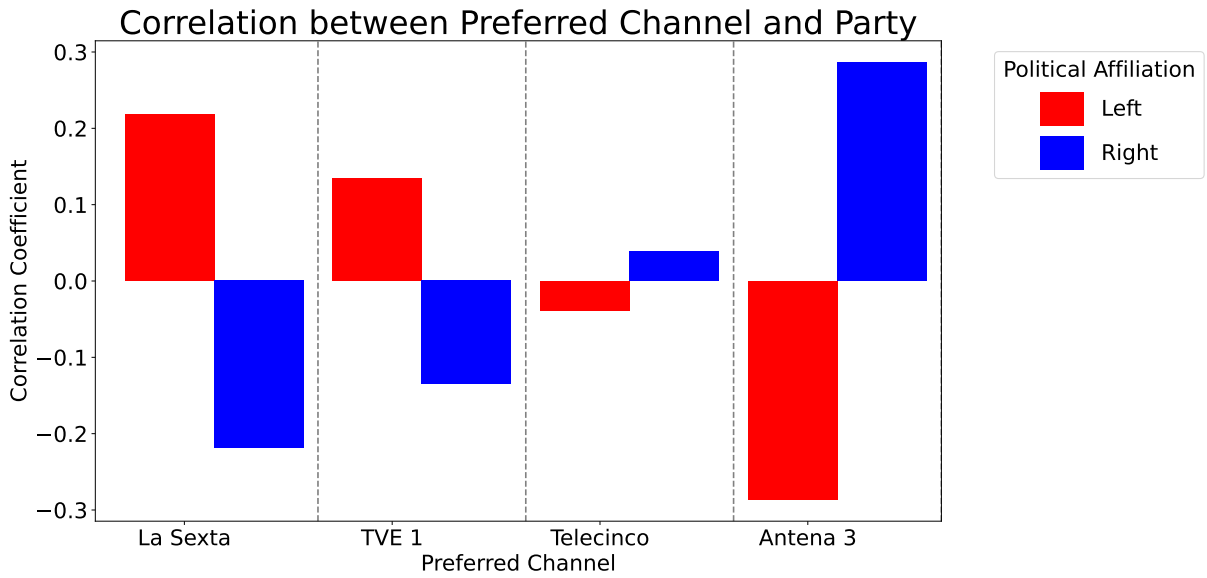

Figure 2: Source: Built using data from CIS's Encuesta Pre-electoral 2023. The survey ask respondents if they watch TV for political content and what is their preferred channel. Bars represent a correlation coefficient between the declared preferred political party (pooling left and right parties) and the most watched channel.

Of course, audiences might be politically different because they sort themselves into the channels they like, because they are persuaded by them or, most likely, a combination of both. It goes beyond the scope of this paper to study these effects.

3 Data and Descriptive Statistics

We have access to a unique dataset that captures both the demand and supply sides of TV news in the Spanish market. For the demand side, we use minute-by-minute viewership data for the four main channels that offer daily TV news programs: TVE, Antena 3, LaSexta, and Telecinco. On the content side, we have daily transcripts that we match to the minute level to see what content people were exposed to. Our data spans from November 2022 to the present, on a daily frequency.

Audience data

We have high-frequency audience data provided by Kantar Media. For our sample period, we can observe the shares of viewers for each channel on a given day and minute. Although we don't have individual data on choices, we do have geographical disaggregation for the 16 Autonomous regions in Spain.[3]

Furthermore, we match this dataset with one from another media provider, Barlovento. The advantage of this data is that it contains manually annotated sections on a minute-by-minute basis, allowing us to split the unstructured text into different segments of the day.

The data is pre-processed by first removing the first 5 minutes of the day where audience shifts are likely driven by inertia and not that much to content characteristics. We also homogenize all the outlets to have the same duration and exclude instances in which regions exhibit 0 market shares.

TV Content

We record daily TV news and use Google Cloud infrastructure to store and process the data with *speech-to-text*. Although visuals play a key role in information transmission, we focus only on text transcripts. A more detailed explanation of the entire downloading pipeline can be seen in Appendix Figure 8.

Agencia EFE

We obtain all news stories provided in Spanish by one of the largest news agencies in the world, Agencia EFE. Due to access limitations, we receive only the title of each story along

[3]The Canary Islands and La Rioja are excluded due to different time zones and zero market shares; respectively.

with a short summary segment. Our sample contains a total of 41K stories.

Survey Data

To understand polarization behavior, we use survey data gathered from the Centro de Investigaciones Sociológicas (CIS). Specifically, we construct the proportion of people intending to vote for right-wing parties in each Autonomous Community each month.

3.1 Political Coverage and Elections

Political power in Spain has historically been dominated by a two-party system, with either the Socialist Party (PSOE) or the Conservative Party (PP) in power. The emergence of the left-wing party Podemos (UP) following the 15M movement marked a significant shift, as the party began to attract a substantial portion of the electorate.[4] Of particular interest is the rise of the far-right party VOX, which made notable gains in the regional elections of 2023, raising the possibility of forming a coalition with the Popular Party (PP) in the regional elections held on May 28, 2023. In response, President Pedro Sánchez decided to bring forward the general elections to June 23, 2023. We refer readers to the Spanish Media Monitor webpage to explore various metrics of coverage across political parties and actors.

Figure 3 shows the average proportion of political mentions in our time period.[5]

Due to sample restrictions, we divide our periods into two. The *pre-campaign* period starts at the beginning of our data collection in December 2022 and extends to the start of the first publicly announced political campaign on May 13, 2023. The *campaign* period covers both regional and general election campaigns and lasts until the day of the general elections, July 17, 2023.

3.2 Tone

Do channels align with their viewers' demand in the political content they offer? In this section, we describe our methodology to classify TV content by political leaning. Traditionally, basic text analysis methods, such as measuring mentions or airtime devoted to political actors, have been used to assess plurality and sometimes interpreted as a positive indicator for providing greater publicity. Similarly, sentiment analysis techniques based on semantics

[4]Relevant for this period of study is the integration of Podemos into the new political party SUMAR. All classification metrics account for this transition, but throughout the text, we refer to UP as either Podemos or SUMAR after its creation.

[5]Political mentions are based on dictionaries that include terms for the main political actors and parties.

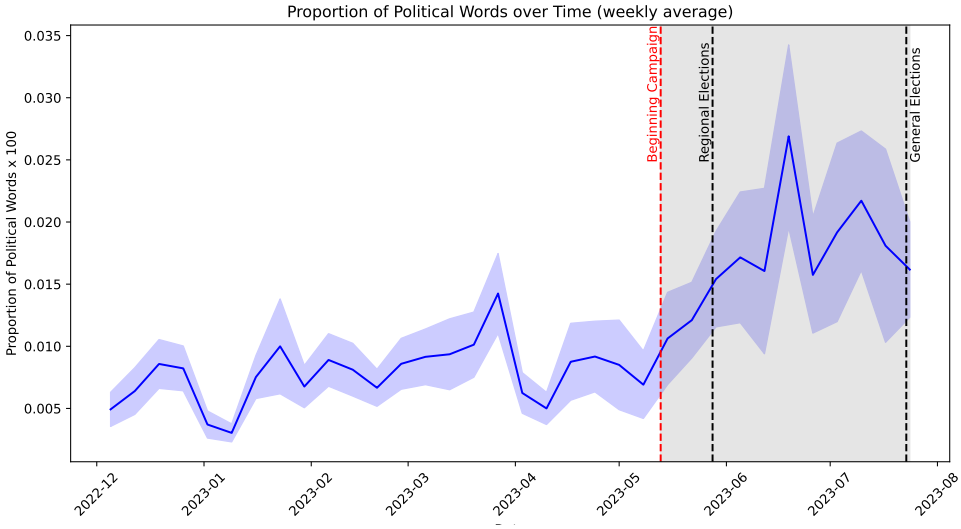

Figure 3: Average daily proportion of political terms relative to overall words with shaded standard deviations. Vertical, dashed lines indicate the date of the regional and general elections, respectively. The shaded area represents the "campaign" period considered.

cannot reliably discern named entities or context. For this reason, we rely on Large Language Models to classify the stories of the day. The use of LLMs as classifiers has gained popularity in recent years for a variety of context such as classification of political stances (Le Mens and Gallego, 2023) and has even been found to archive higher precision and accuracy scores for ideological classification when compared to human annotators (Törnberg, 2023).

Specifically, we use the latest version of ChatGPT-4, feeding it all our political stories and asking it to classify the tone associated with each political party. Notably, we distinguish between positive and negative tones toward the four political parties considered, and we provide a flexible query that allows the classifier to be neutral if the content is ambiguous. More details about the prompt and classification results can be found on Appendix 8.2

The average tone for each channel-party combination after classification is shown in Figure 4. Both Figures 2 and 4 look similar and indicate that viewers self-select into channels that better match their political affiliation—a finding consistent with previous works (Gentzkow et al., 2011) that highlights market equilibrium.

4 Preferences for Politics

4.1 Political Tone and Electoral Campaign

The analysis in the previous sections indicates a general distaste for hard news but does not allow us to comment on tolerance for different political parties. In this section, we explore

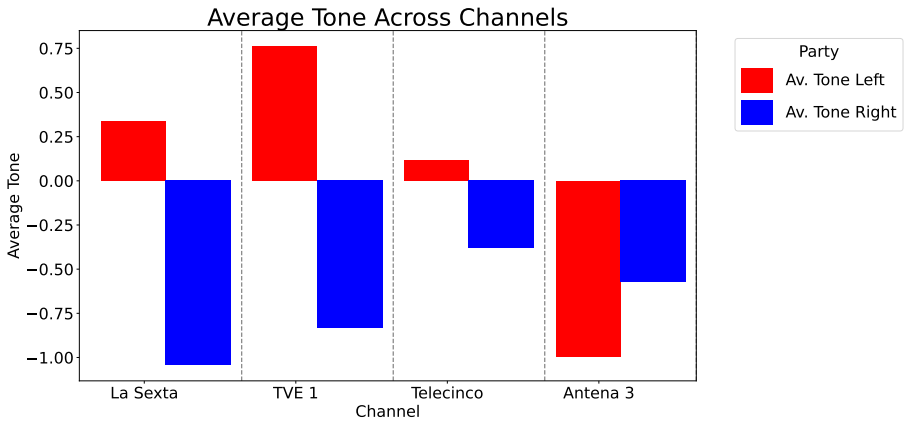

Figure 4: Average tone for each channel-party as classified by Chat GPT 4 from the whole sample period.

whether there has been asymmetric distaste for different political parties in response to the campaign period. We begin by detailing how channels have varied their tone toward the parties during this period (Wire, 2023).

We decompose relative tone before and during campaign periods in Figure 5, relative to the number of minutes covered on national politics. The left-leaning channel, LaSexta, significantly reinforced its left position by increasing the positive tone on left-wing parties and decreasing it on both right-wing parties. All other outlets opted for a moderating effect, reducing their mean relative tone.

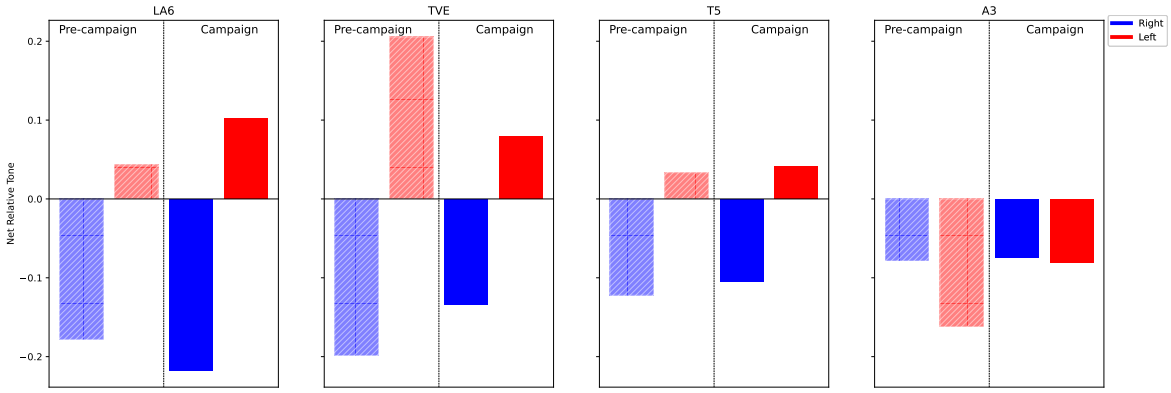

Figure 5: The figure shows the relative tone calculated as the number of minutes pro right minus the number of minutes anti right relative to the total number of minutes devoted to national politics. The vertical dashed line shows results pre and during campaign periods, respectively.

5 Market set up

Our market set up differs from the classical demand estimation problem as there is absence of prices and channels differentiate themselves varying product characteristics. We estimate

demand using a mixed logit model (Berry, 1994) [6].

An individual i a given market t (region-day), chooses an outlet j based on the following indirect utility [7]:

$$U_{ijt} = \underbrace{\sum_k x^k_{jd(t)}\alpha^k + \xi_{jt}}_{\delta_{jt}} + \underbrace{\sum_k x^k_{jd(t)}\left(\sigma^k \nu^k_i + \pi^k y_{it}\right)}_{\mu_{ijt}} + \epsilon_{ijt} \tag{1}$$

where x^k_{jt} represents the kth characteristic on channel j and day $d(t)$ and σ^k is a shift on characteristic k mean preferences according to unobserved preferences $\nu^k_i \sim N(0,1)$. For content characteristics we consider the proportion of positive and negative minutes devoted to each party and the relative amount of minutes spent in political issues to control for a seek of variety in content other than politics. We also introduce as demographics, y_{it}. We use survey data to calculate the proportion of right wing intention to vote in a given region and month. Thus the parameters π^k enable us to capture polarization behavior where more right-wing viewers have asymmetric tastes for content on their own and the opposed party tone.

Our preferred specification decomposes the unobserved characteristics into $\xi_{jt} = \xi_j + \xi_{dow} + \Delta\xi_{jt}$, where we include product dummies that account for unobserved quality factors and day of the week dummies to control for the seasonality in our audience data. We also cluster the standard errors at the region level.

The outside option is modeled in terms of *potential* audience. Formally, the observed share of people consuming a channel is $s_{jt} \equiv \frac{q_{jt}}{L_{r(t)}}$where q_{jt} is the total number of viewers watching channel j in region-day t and $L_{r(t)}$ is the potential of viewers having access to TV on region $r(t)$; defined by Kantar media as the population that has access to television in a given region.

Market shares of channel j on region-day t can be approximated by a sum over the individuals in a given market, $\mathcal{I}_t$ as :

$$s_{jt} = \sum_{i \in \mathcal{I}_t} w_{it} \frac{exp(\delta_{jt} + \mu_{ijt})}{1 + \sum_{l \neq j} exp(\delta_{lt} + \mu_{ilt})} \tag{2}$$

where w_{it} are survey weights the control for the different relative weight of each region.

The nature of the horizontal differentiation game makes product characteristics correlated with unobserved local shocks to viewers' preferences. We explain in detail our instrumental variables strategy to mitigate this concern in the next section.

[6] The model is estimated under the *pyblp* package (Conlon and Gortmaker, 2020).

[7] See digression 8.3 on the Appendix for a re-interpretation of the discrete choices.

5.1 Content endogeneity and News Shocks

Under preferences in equation 1, content characteristics are an equilibrium outcome of the differentiation game and therefore we expect the unobserved vertical quality component, ξ_{jt} to be correlated with them. In this set up, endogeneity arises because local shocks to viewers' preferences are taken into account by the outlets before broadcasting their content. Even though channels cannot adapt content at the region level, Figure 10 shows the political density in the left right spectrum of each channels' audience. Consistent with survey evidence outlined above, channels are aware of their audience composition and thus internalize political shocks that might affect viewer tastes in a given time period. To address this endogeneity concern, we instrument for each of the k linear coefficients, as well as the non-linear ones. To instrument for the former, we target the news production of each outlet.

Suppose that each day contains stories indexed by $s \in S_d$. A story can be about party $p(s) \in \{L, R, \emptyset\}$ and have a tone $\tau(s) \in \{-1, 0, 1\}$ [8]. We define the total number of positive and negative stories available on a given day as the total sum of the tone for $party \in \mathcal{P}$ in the News Agency (i.e., Agencia EFE) dataset. Specifically:

$$
\begin{aligned}
pos(party)_d^{NEWS} &= \frac{1}{|S_d|} \sum_{s \in S_d} \Big(\mathbb{1}\{\tau(s) = 1\} \times \mathbb{1}\{p(s) = party\} \Big) \qquad \forall party \in \{L, R\} \\
neg(party)_d^{NEWS} &= \frac{1}{|S_d|} \sum_{s \in S_d} \Big(\mathbb{1}\{\tau(s) = -1\} \times \mathbb{1}\{p(s) = party\} \Big) \qquad \forall party \in \{L, R\} \\
political_d^{NEWS} &= \frac{1}{|S_d|} \sum_{s \in S_d} \Big(\mathbb{1}\{p(s) \neq \emptyset\} \Big)
\end{aligned} \tag{3}
$$

We take the equilibrium of the differentiation game as fixed in the short run (i.e., average results from 4). This notion of short-run equilibrium is crucial to justify why shocks have product-level variation, even though we take a common set of political stories for all of them. In essence, it is the cost of maintaining the editorial position that makes the day's composition affect outlets differently. This allows us to place channels on a left-right scale given the *intrinsic ideology* in the differentiation game, as illustrated in Figure 6. This is crucial as it is the outlet's position *together* with the daily composition of stories $(pos(party)_d^{NEWS}, neg(party)_d^{NEWS}, political_d^{NEWS})$ that gives variation in news production on both the outlet and market dimensions.

In order to illustrate how the inputs interact with the ideological news production of each outlet, we pool (for simplicity) left channels (La Sexta and TVE), middle (Telecinco) and right (Antena 3) and run regressions of the form:

[8] Both political but neutral stories and non-political stories are grouped into a tone of 0.

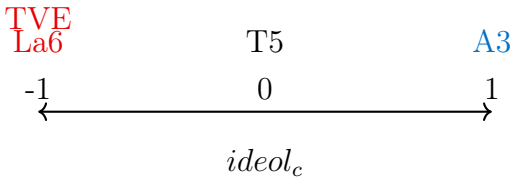

Figure 6: Illustration of the channel's equilibrium position in the political spectrum according to their average reported tone on left and right channels.

$$\begin{aligned} pos(p)_{jt}^{CH} = &\sum_{p\in\mathcal{P}}\sum_{j'\in\mathcal{J}} \left(d(j')_j \times pos(p)_d^{NEWS}\right) \beta_j^{(p,+)} + \\ &+ \sum_{p\in\mathcal{P}}\sum_{j'\in\mathcal{J}} \left(d(j')_j \times neg(p)_d^{NEWS}\right) \beta_j^{(p,-)} + \epsilon_{jt} \end{aligned} \tag{4}$$

$$\begin{aligned} neg(p)_{jt}^{CH} = &\sum_{p\in\mathcal{P}}\sum_{j'\in\mathcal{J}} \left(d(j')_j \times pos(p)_d^{NEWS}\right) \gamma_j^{(p,+)} + \\ &+ \sum_{p\in\mathcal{P}}\sum_{j'\in\mathcal{J}} \left(d(j')_j \times neg(p)_d^{NEWS}\right) \gamma_j^{(p,-)} + \epsilon_{jt} \end{aligned} \tag{5}$$

$$political_{jt}^{CH} = \sum_{j'\in\mathcal{J}} \left(d(j')_j \times political_d^{NEWS}\right) \phi_j + \epsilon_{jt} \tag{6}$$

where $(pos(p)_{jt}^{CH}, neg(p)_{jt}^{CH}, political_{jt}^{CH})$ are the analogous to 3 but for each outlet. In this case everything is normalized by the total number of minutes [9]

A positive $\beta_{ch}^{R,+}$ coefficient indicates that given an extra positive story about the right wing party is available, outlet ch increases the number of positive stories about that party by $\beta_{ch}^{R,+}$, ceteris paribus.

The results of the estimation are shown in table **??** on the Appendix. For illustration purposes, we show the predicted production of political stories under different day compositions in figure ??. We pool parties into left and right groups and simulate 4 days with the positive/negative combinations for the right and left parties and plot the predicted channel's daily composition.

Bars represent the difference between the predicted positive and negative stories for a given party. The top left panel shows that on days highly unfavorable for the left, it is the right channel that presents a more negative scenario for that party. Conversely, days with dominant positive left stories (bottom left graph) only look favorable from the left channels perspective. This illustrates that, even though the shock of news is common to all channels, their intrinsic ideology makes them asymmetrically constrained in their news production

[9]Even though the classification of ChatGPT is fed with the splited stories, we then weight back those stories into the minutes that they span to obtain a measure of how many positive or negative minutes are told by a given outlet in a given day.

and can be used as an instrument. The exclusion restriction implies that audience does not ex-ante react to different composition of the day. For instance, if viewers knew that a scandal happened and they switched to their preferred outlet to watch it, that would violate our identification assumption, Although this cannot be tested directly, we show that using the minute 0 audience (i.e before any content has been shown already), we do not find asymmetric audience gains across outlets based on the composition of the day. Results are shown in appendix table 3.

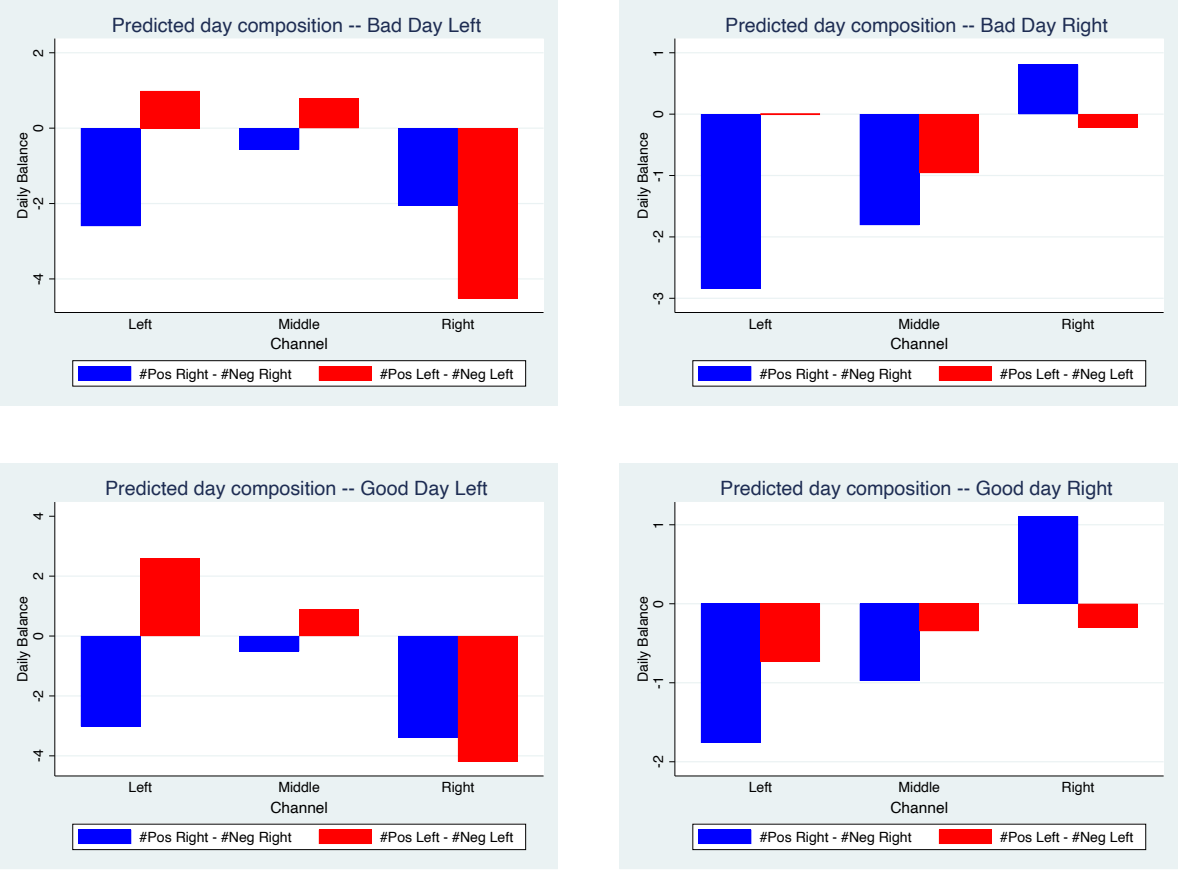

Figure 7: Predicted number of positive minus negative minutes for each outlet and party. The top left (right) plots show days where 4 negative stories for the left(right) party are available and 1 stories of the other type for the rest. The bottom graphs show the analogous scenario with positive stories. The horizontal axis represents the channel classified by its intrinsic ideology into left (La6 and TVE), Middle(Telecinco) and Right (Antena3).

For the k non linear coefficients that represent the variance of the individual heterogeneity we follow Gandhi and Houde (2019) and target predicted product differentiation instruments of the form $\left(\hat{x}^k_{jt} - \sum_{l \neq j} \hat{x}^k_{lt}\right)^2$, where $\hat{x}^k_{jt}$ is the predicted kth characteristic from the first stage of the instrumental variable regression 3. To target the demographic parameters, we interact the mean proportion of right votes with those instruments: $\bar{y}_t \left(\hat{x}^k_{jt} - \sum_{l \neq j} \hat{x}^k_{lt}\right)^2$.

6 Results

We present results of the BLP estimation for the pre-campaign and campaign periods in table 1. We show the preferred specification under outlet and day of the week fixed effects with standard errors clustered at the region level.

The pre-campaign period estimation shows a significant, positive average taste for negative stories on the right party with significance dispersion ($\sigma_{\text{neg\_right}}$) in the population. Importantly, this dispersion is not explained by ideology heterogeneity (π coefficients).

The campaign period estimates showed at the bottom of the table show a somewhat different set up. There is a significant positive taste for positive stories on the left and negative stories on the right parties. This fact, together with the negative taste on positive right and negative left content would indicate an average preference towards the left leaning parties. The coefficients with the interaction with ideology further indicate a result in line with a polarization behavior: more right wing audiences demand content that is more favorable towards their own party and more negative towards the opposition. The fact that this behavior is exhibited only during the political campaign period and not before links to several results on the political information acquisition.

The taste for national politics overal is negative with right wing viewers having a higher demand for it.

7 Incorporating the supply side

Work in progress

In this section we incorporate a supply side into the previous market. Outlets are pure profit maximizers and focus on maximizing their share of audience. However, they are affected by input shocks that limit their news production. In particular, we model profits for a given channel j on a given day d as :

$$\Pi_{jd} = \sum_{r} s_{jrd}(\boldsymbol{x}_{jd}) - (\lambda_j \mathcal{C}(\boldsymbol{x}_{jd}, \boldsymbol{z}_d) - \phi_j \times ideol_j) + \boldsymbol{\eta}_{jd}\boldsymbol{x}_{jd} \tag{7}$$

where the first term is the sum of the market audience shares over all regions. The second term within parenthesis represent the news' production costs. The parameter λ_j captures the marginal cost estimate of characteristics as a function of the inputs available on day d, $\boldsymbol{z}_d$. The cost function $\mathcal{C}(\cdot)$ is assumed to be decreasing in $\boldsymbol{z}_d$ and convex in $\boldsymbol{x}_{jd}$. Intuitively, producing more stories about a political party becomes cheaper on days where there are more stories about it. However, as explained in previous sections, channels face asymmetric costs

that depend on their ideological stand on the game. This is modeled through a reduction in costs with the weight given to departing from the position parameterized by ϕ_j. The last term introduces content specific shocks.

Coefficient	Parameter	Estimate	Std. Error
	Pre-campaign		
pos_left	$\sigma_{\text{pos\_left}}$	6.59	(24.98)
pos_right	$\sigma_{\text{pos\_right}}$	23.29	(21.09)
neg_left	$\sigma_{\text{neg\_left}}$	5.13	(8.21)
neg_right	$\sigma_{\text{neg\_right}}$	25.11***	(5.42)
political	$\sigma_{\text{political}}$	3.20	(8.53)
mean_right × pos_left	$\pi_{\text{pos\_left}}$	21.28	(16.80)
mean_right ×pos_right	$\pi_{\text{pos\_right}}$	37.82	(43.96)
mean_right ×neg_left	$\pi_{\text{neg\_left}}$	-47.84	(33.99)
mean_right ×neg_right	$\pi_{\text{neg\_right}}$	-64.89	(49.25)
mean_right × political	$\pi_{\text{political}}$	-13.94	(12.83)
pos_left	$\beta_{\text{pos\_left}}$	-10.71	(7.74)
pos_right	$\beta_{\text{pos\_right}}$	-18.76	(23.73)
neg_left	$\beta_{\text{neg\_left}}$	14.57	(12.80)
neg_right	$\beta_{\text{neg\_right}}$	43.31***	(16.62)
political	$\beta_{\text{political}}$	5.02	(5.70)
	Campaign		
pos_left	$\sigma_{\text{pos\_left}}$	0.89	(192.62)
pos_right	$\sigma_{\text{pos\_right}}$	31.42	(42.11)
neg_left	$\sigma_{\text{neg\_left}}$	0.50	(223.28)
neg_right	$\sigma_{\text{neg\_right}}$	0.47	(176.24)
political	$\sigma_{\text{political}}$	2.91**	(1.21)
mean_right × pos_left	$\pi_{\text{pos\_left}}$	-632.49***	(217.31)
mean_right × pos_right	$\pi_{\text{pos\_right}}$	461.89***	(137.64)
mean_right × neg_left	$\pi_{\text{neg\_left}}$	377.37***	(138.78)
mean_right × neg_right	$\pi_{\text{neg\_right}}$	-409.18***	(127.43)
mean_right × political	$\pi_{\text{political}}$	25.05**	(12.14)
pos_left	$\beta_{\text{pos\_left}}$	194.60***	(52.05)
pos_right	$\beta_{\text{pos\_right}}$	-174.47**	(75.41)
neg_left	$\beta_{\text{neg\_left}}$	-134.62***	(45.20)
neg_right	$\beta_{\text{neg\_right}}$	132.97***	(33.80)
political	$\beta_{\text{political}}$	-9.94*	(5.39)

Table 1: BLP Estimation Results with Standard Errors

The table shows the results of the BLP estimation of model 1. The estimations are divided in the pre-campaign and campaign period. Both day of the week and outlet fixed effects are included. Standard errors are clustered at the region level.

8 Appendix

8.1 Figures

Pipeline:

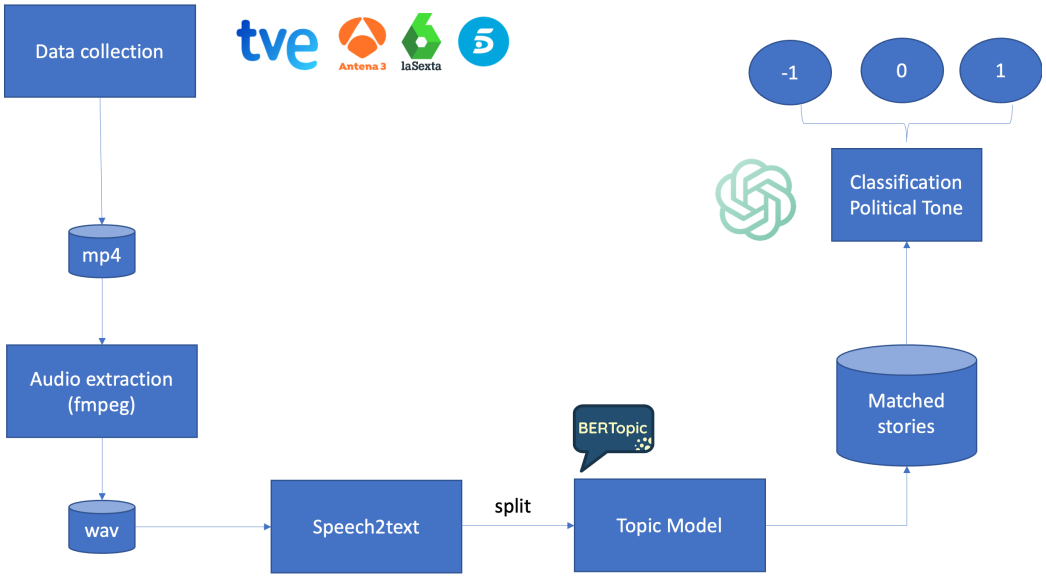

Figure 8: Pipeline for the text downloading. First videos are downloaded daily from the main TV channels. Google engine is used to convert the audio to text. We then split the stories by minute and user BERTopic to classify and match them. Finally, ChatGPT4 is used to classify political tone.

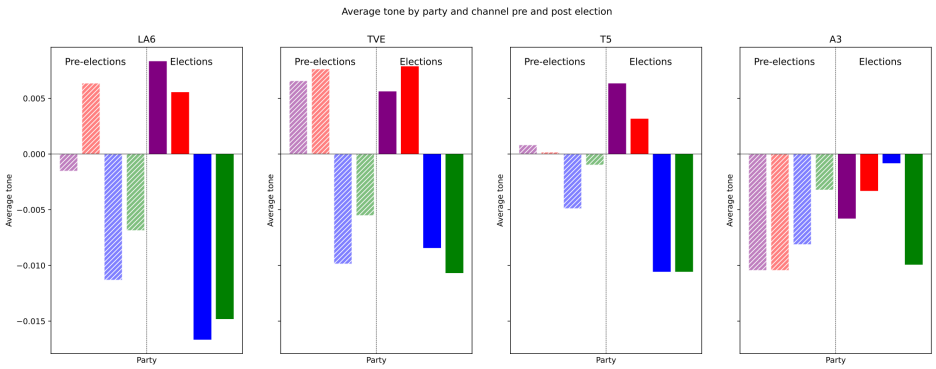

Figure 9: Average tone for each party and channel pre and during campaign periods

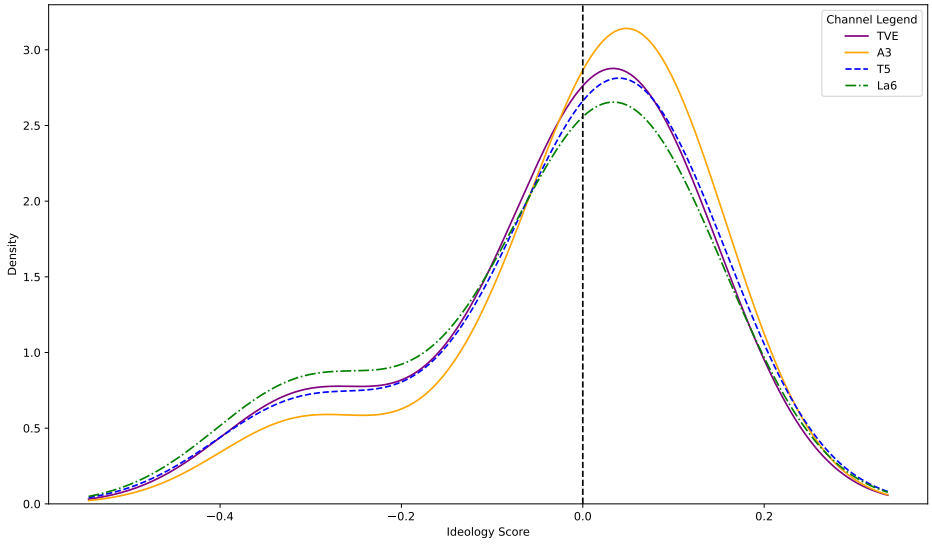

Figure 10: Estimated density of channels' audience ideology. The figure shows a kernel density estimate on the ideology score constructed using survey daa and weighted by channels' share of audience for each autonomous region.

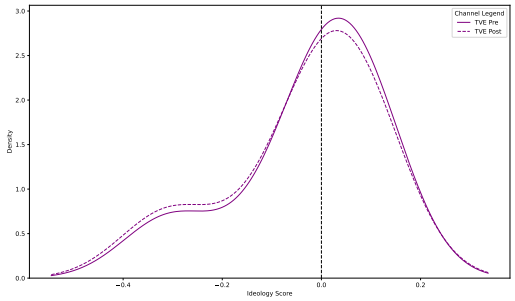

Figure 11: Estimated densities pre and during campaign for TVE.

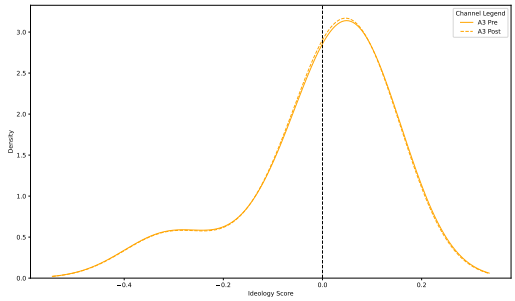

Figure 12: Estimated densities pre and during campaign for Antena 3.

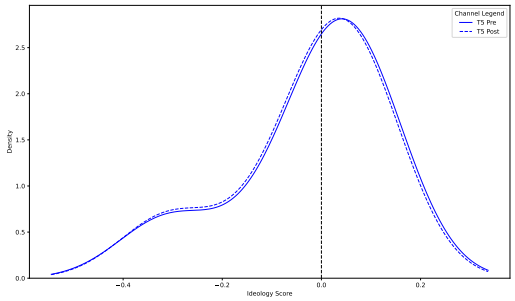

Figure 13: Estimated densities pre and during campaign for Telecinco.

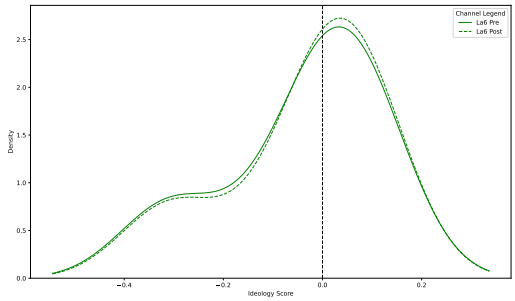

Figure 14: Estimated densities pre and during campaign for La Sexta.

8.2 Chat GPT ideology classification

In this section we summarize the usage of ChatGPT as a text classifier for political tone. We detail the prompt and specification details used for the text classification together with final results.

To reduce both computational and monetary costs, we first filter our split stories using a simple dictionary based approach into those that might contain any relevant political information. Table 2 shows the key terms used to filter the political stories. After the match, we obtain a final number of 15406 political stories that we feed into the chat GPT classifier.

Political	**PP**	**PSOE**	**SUMAR/UP**	**VOX**
política	pp	psoe	unidas podemos	vox
democracia	partido popular	partido socialista	podemos	abascal
partido político	feijoo	sanchez	ione belarra	de los monteros
gobierno	alberto nunez feijoo	federico buyolo garcia	pablo iglesias	macarena olona
elecciones	ayuso	maria jesus montero	yolanda diaz	ortega smith
votación	cuca gamarra	carmen calvo	irene montero	rocio monasterio
constitución	pablo casado	jose luis abalos	alberto garzon	ignacio garriga
legislación	esperanza aguirre	felix bolanos	iona errejon	jose alcaraz
senado	ana pastor	francina armengol	monica garcia	herminio campillo
congreso	pilar barreiro	sanchez mato	jaume asens	zambrano garcia
dictadura	rafael hernando	margarita robles	noelia vera	luis gestoso
soberanía	alvarez de toledo	marlaska	raul camargo	
estado	javier maroto	jose manuel albares	lopez de uralde	
ciudadanía		isabel rodriguez	rosa martinez	
derechos				
libertades				

Table 2: The table shows the political words included to filter the stories into national politics. We included both general terms that refer to politics as well as party specific terms.

We use OpenAI PI research using model *gpt-4-0125-preview* to build queries of the form:

Prompt

Analyze the sentiment of the following news article with respect to the political parties (and their members) in Spain: PP, Podemos/Sumar, PSOE, VOX. Only use numeric values from the set [-1, -0.5, 0, 0.5, 1].
Evaluate the sentiment towards each party with a number between -1 and 1, where -1 indicates an extremely negative perception, 0 indicates neutrality or irrelevance for the party, and 1 indicates an extremely positive perception.
Consider only the values -1, -0.5, 0, 0.5, and 1.
Base your evaluation solely on the explicit content of the news article. If the article does not mention or imply any sentiment towards a party, assign a 0 to that party.
The format must always be a list `[PP , PSOE , UP , VOX ]` where `X` represents the numeric sentiment value.

Due to the stochasticity in LLMs predictions, good practices recommend to run the classification multiple times and avere out the results (Törnberg, 2023). Financial costs, however, impede us from doing the whole approach multiple times but we show results on stability below on a subset of stories.

Results of the final classification for the non neutral stories are shown in figure 15. Each bar represents the percentage of stories of that given sentiment associated to each political party. We can see that the classification reserved the extreme values 1 and -1 for few stories and focused on the -0.5 and 0.5 values.

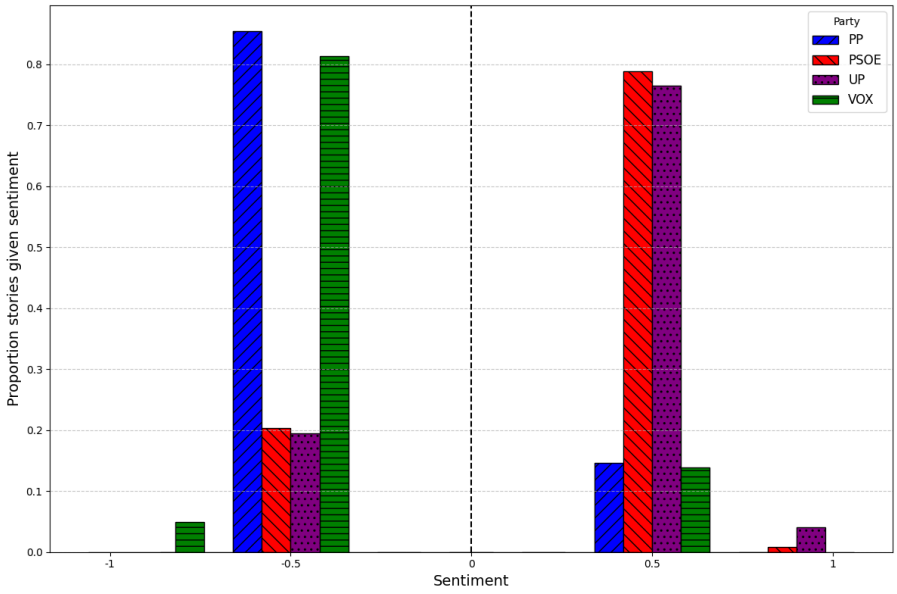

Figure 15: The figure shows the percentage of stories to each political party for a given sentiment. We exclude neutral stories.

Statistic	PP	PSOE	VOX	UP
Mean	-0.013759	0.106458	-0.053395	0.024250
Standard Error	0.003166	0.003582	0.001345	0.001970

Note: The table shows the mean and standard error for 100 rounds of Chat GPT classification of political with 40 random political stories.

8.3 Digression: Minute-by-minute decisions vs single discrete choices.

The discrete choice model outlined in section 5 makes individuals choose a single option (channel or outside option) per day. We rationalize this behavior to the fact that individuals cannot possibly observe simultaneity all the channel characteristics to then pick their favorite one, but rather discover them by zapping, which implicitly implies audience shifts. Suppose that individual do high-frequency(i.e minute-byminute) decisions on whether to stay in a channel or switch to another option. Specifically, denote by $a_{it} \in \{\mathcal{J} \cup \emptyset\}$ the decision of individual i on minute t. Individuals start the day in a predefined option and at each minute decide whether to stay in that channel or move to another option with transition probabilities:

$$Pr(a_{it} = k | a_{it-1} = j) = \begin{cases} Pr(f_i(x_{j,t-1}) + \epsilon_{ijt} \geq \bar{u}_i) & if \quad k = j \\ \frac{1}{|\mathcal{J}|} & if \quad k \neq j \end{cases} \tag{8}$$

that is, individuals remain in the same channel if the utility that the derive from the content they have watched is larger than some reservation utility and else switch to a different alternative with uniform probability. The expected audience for a channel in a given minute is :

$$\mathbb{E}(q_{jt}) = L \sum_i \sum_{k \in \mathcal{J}} Pr\left(a_{it} = j | a_{i,t-1}\right) Pr\left(a_{i,t-1} = k\right) \tag{9}$$

from where the expected audience of a given day is just $\frac{1}{T}\sum_t \mathbb{E}(q_{jt})$. Therefore the market shares in equation 2 can be interpreted as time share allocations from a model where individuals switch across channels in a dynamic fashion throughout the day.

8.4 Instrument validity

The instrumental variable approach outlined in section **??** relies on the assumption that the supply shifters are uncorrelated to demand. This assumption would be violated if viewers exhibit some anticipation behavior. For example, individuals might get information about what has happened on a given day and tune into the channel they prefer accordingly. Although this cannot be tested directly, we propose a simple test of this behavior in a

	$ln\left(\frac{q_{crd}}{L_r - Q_{rd}}\right)$			
	(1)	(2)	(3)	(4)
pos_right_news	-1.1830***	-1.2970***	-0.9093***	-0.9572***
	(0.2379)	(0.2059)	(0.2365)	(0.2832)
pos_left_news	0.2434	0.8154***	0.6709***	0.6657***
	(0.2172)	(0.1709)	(0.2461)	(0.2339)
neg_right_news	-0.6172***	-0.7594***	-1.0197***	-0.5473**
	(0.2177)	(0.1771)	(0.2712)	(0.2548)
neg_left_news	1.5192***	1.7982***	1.0508***	1.0127***
	(0.2582)	(0.2026)	(0.2770)	(0.2861)
political_news	-0.5071**	-0.5051***	0.0565	-0.3992*
	(0.2029)	(0.1723)	(0.2213)	(0.2388)
_cons	-3.3131***	-2.7860***	-3.6418***	-4.1093***
	(0.0650)	(0.0546)	(0.0671)	(0.0724)
N	2231	2555	2001	2381

Standard errors in parentheses

*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$

Table 3: The table shows the results from the estimation of 10 conditional on minute 0 audience. Each column represents the regression on channels TVE, A3, Telecinco and La Sexta ; respectively. Day of the week and region fixed effects are included.

reduced form approach. We want to see if the minute 0 audience is correlated to the political composition of the day. By considering the initial audience, we make sure that channel's content has not play a role yet. Specifically, we run channel by channel regressions of the form:

$$ln\left(\frac{q_{crd}}{L_r - Q_{rd}}\right) = \beta_0 + \sum_{p\in\{L,R\}} \left(\beta_p pos(p)_d^{NEWS} + \beta_p neg(p)_d^{NEWS}\right) + \beta_{pol} pol_d^{NEWS} + \gamma_{dow} + \gamma_r + \epsilon_{crd} \tag{10}$$

Table 3 show the estimated coefficients from equation 10 where each column corresponds to a separate outlet regression and we condition on our minute zero audience. Equation 10 is the linear analogue of a logit estimation. The consistent sign of the coefficients across the different covariates indicates that viewers might not strategically choose across the outlets as a function of what the stories of the day were. This, however, doesn't preclude strategic substitution towards the outside option.

8.5 Alternative methods for segment splitting in unstructured text

We explain here different unsupervised text splitting methods that were tried to split the segments of the day. To the best of our knowledge, these methods have not been applied before and can be easily extended to any TV news set up. The advantage of these techniques is that they provide an unsupervised way to split unstructured text into stories that is precise up to the second level, thus overcoming the problem of the manually annotated labels which goes at the minute level. There is, however, computational or financial costs in some of them that impeded us to use them for our whole dataset.

Image recognition

Outlets typically segment their sections by means of captions where they introduce headers for the upcoming story. Exploiting our video dataset, we designed an unsupervised image recognition algorithm that tracks the appearance of those new segments and produces a set of times that serve as text splitters. Although precise, the disadvantage of this method is that it remains computationally intensive as videos need to be segmented and then processed into the algorithm. Computational cost can be reduced by focusing on the lower bottom of the screen only (figure ??), which is the area where the output is expected to appear.

Figure 16: Example of a image with a caption that delimits the beginning of a new section. The highlighted area shows the bottom of the image where image recognition can be applied to find such appearances.

Speaker diarization

A less computationally expensive alternative consist on using *speaker diarization* on the wav files. After transforming the mp4 file to audio using *ffmpeg*, we use Google Cloud diarization tool to find the different speakers in an audio file. The most common (or most two common if it is a weekend) speakers overall corresponds to the presenter of the news. After allowing a flexible specification, one can identify break points by cheeking the seconds where the presenter comes back into scene making sure she is speaking long enough so that a new segment is being introduced. Figure 17 illustrates this procedure and the comparison with

the manually annotated labels for an example day-channel.

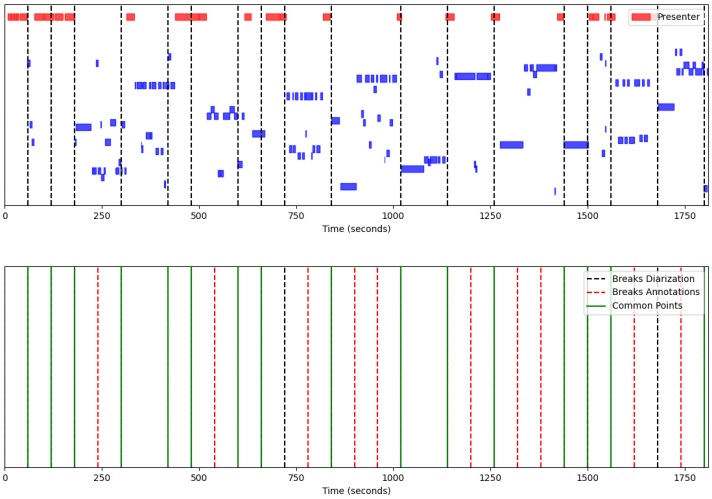

Figure 17: The top figure shows the timeline for the presenter audio (red) vs other audios recognized by *speech2text* in a wav file for the 15th January 2023 in La Sexta. Vertical, black, dashed lines represent the predicted splits based on the diarization. The bottom figure combines these splits with the ones that come from the manually annotated figures. Red, dashed bars correspond to the breaks that come from the manually annotated GECA dataset. Green bars represent breaks where both the speaker diarization and the manual annotation coincide on a break.

References

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Centro de Estudios Murciano de Opinión Pública (CEMOP) (2024). La polarización afectiva crece en españa un 30

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with pyblp. *The RAND Journal of Economics*, 51(4):1108–1161.

Durante, R. and Knight, B. (2012). Partisan control, media bias, and viewer responses: Evidence from berlusconi's italy. *Journal of the European Economic Association*, 10(3):451–481.

Edelman (2023). Edelman trust barometer 2023: Global report.

Enikolopov, R., Petrova, M., and Zhuravskaya, E. (2011). Media and political persuasion: Evidence from russia. *American Economic Review*, 101(7):3253–85.

Gambaro, M., Larcinese, V., Puglisi, R., and Snyder Jr, J. M. (2021). The revealed demand for hard vs. soft news: Evidence from italian tv viewership. *National Bureau of Economic Research*, Working Paper 29020.

Gandhi, A. and Houde, J.-F. (2019). Measuring substitution patterns in differentiated products industries. Technical Report 26375, National Bureau of Economic Research.

Gentzkow, M., Shapiro, J. M., and Sinkinson, M. (2011). Competition and ideological diversity: Historical evidence from us newspapers. *American Economic Review*, 101(7):1927–1958.

Larcinese, V. (2003). The instrumental voter goes to the news-agent: Demand for information, election closeness, and the media.

Le Mens, G. and Gallego, A. (2023). Positioning political texts with large language models by asking and averaging. *arXiv e-prints*, page arXiv:2311.16639.

Parliament, E. (2024). Audiovisual and media policy.

Patterson, T. E. (2000). Doing well and doing good. (RWP01-001).

Peterson, E., Goel, S., and Iyengar, S. (2017). Echo chambers and partisan polarization: Evidence from the 2016 presidential campaign. In *Proceedings of the Annual Meeting of the American Political Science Association*.

Prior, M. (2009). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, 71(3):893–908.

Puglisi, R. and Snyder, J. M. (2015). Chapter 15 - empirical studies of media bias. 1:647–667.

Törnberg, P. (2023). Chatgpt-4 outperforms experts and crowd workers in annotating political twitter messages with zero-shot learning.

Wire, A. N. (2023). Ties with far-right hinder spain's popular party forming a government despite election win. *The Independent*.