The Impact of Political Campaigns on Demand for Partisan News

Luis Menéndez *

March 21, 2025

Abstract

How do individuals acquire political information during election campaigns? This paper examines the demand for political news using a unique dataset that combines audience viewership data with the full text of news stories from Spanish TV channels. I estimate a structural random coefficients demand model to capture heterogeneity in political preferences, incorporating demographic differences. A key challenge in measuring demand for political content is the endogeneity of news supply. To address this, I propose a novel identification strategy that exploits exogenous variation in the daily news landscape. By classifying all stories produced in Spain on a given day, I construct a set of potential inputs available to TV channels. Since channels have established political stances and face short-run costs in deviating from them, fluctuations in the political composition of available stories constrain left- and right-leaning outlets asymmetrically. Using Large Language Models, I categorize the tone of each story with respect to political parties. My findings show that political campaigns significantly alter news consumption patterns. While there is no clear partisan asymmetry in demand before campaigns, polarization emerges once the campaign begins. Right-leaning viewers increasingly seek positive coverage of their own party and negative coverage of the opposition. These results highlight the role of campaign dynamics in shaping media consumption and contribute to the broader literature on political media bias and voter behavior.

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

1 Introduction

Media plays a crucial role in disseminating information and can influence voting outcomes. Consequently, policymakers often regulate media markets to ensure informational plurality. One common approach is enforcing proportional airtime rules during political campaigns, a policy implemented in several countries.

However, assessing the effectiveness of these regulations requires understanding how viewers respond to the content they consume. Previous research has demonstrated media's influence on voting behavior during political campaigns (Enikolopov et al., 2011). This issue is particularly relevant in Spain, where political polarization ranks among the highest in Europe, according to recent surveys (Edelman, 2023). Notably, affective polarization has increased significantly during our sampling period, raising the question of whether citizens' political news consumption patterns reflect similar trends.

In this paper, I use a unique dataset that combines Spanish TV news demand from audimeter data with the "product characteristics" of news channels to estimate preferences for political content through a revealed preferences approach; thereby addressing some of the well-known limitations of survey-based studies (Prior, 2009).

My analysis focuses on how TV news demand changes during political campaign periods. Previous research has shown that the rise of soft news has contributed to declining interest in political news (Patterson, 2000). Consistent with Gambaro et al. (2021), my findings suggest that periods in which political content is mixed with other types of programming encourage greater audience engagement with politics.

I then investigate the extent of polarized news consumption. Using different machine learning methods, I desing a pipeline that extracts the daily news videos, converts the audio into text and splits the different stories of the day. This methodology allows me to build comparative statics on tone and duration of the same story treated by different outlets on a given day. Using text analysis techniques, I control for the political stance and duration of news stories (Puglisi and Snyder, 2015). Specifically, I make use of Large Language Models (LLMs) by inputting all stories into ChatGPT-4, prompting it to identify the tone various political parties. The tone analysis across outlets aligns with viewers' self-reported preferences in survey data, allowing us to classify outlets as left-, right-, or center-leaning based on the parties they most support.

To isolate content choices from demand- or supply-driven factors (Gentzkow et al., 2011), I assume that news channels engage in a purely horizontal competition, adjusting their content to attract viewers. I then estimate a discrete choice model (Berry, 1994), in which individuals select their preferred channel based on the political content offered.

In this framework, the political tone offered by news outlets emerges as an equilibrium outcome, making it endogenous in the demand estimation. To address this endogeneity, I exploit random variation in the inputs available to channels. Specifically, I classify all daily news stories produced in Spain during the sample period using data from a major news provider and treat them as the potential story pool for TV outlets. Since channels have established political stances within the differentiation game and face short-run costs in deviating from them, daily fluctuations in the political composition of available stories constrain left- and right-leaning channels differently.

My findings align with previous research on polarized news consumption in the U.S. (Peterson et al., 2017). In the pre-campaign period, there is no systematic asymmetry in political content demand between right- and left-leaning audiences, even when controlling for voting intentions or channel-specific preferences. However, during political campaigns, I find evidence of affective polarization: right-leaning viewers increasingly seek negative coverage of the opposing left-wing party while favoring more positive coverage of their own. This shift in demand coincides with a homogenization of political stances across outlets. Moreover, my results suggest a partisan divide in the overall preference for political content.

This study contributes to the literature on political preferences in media consumption by leveraging exogenous variation in the news landscape to identify political preferences. Notably, my approach relies on unsupervised classification techniques that streamline political categorization. To the best of my knowledge, this is the first paper to examine preferences for political content during campaign periods while explicitly addressing the endogeneity in content supply.

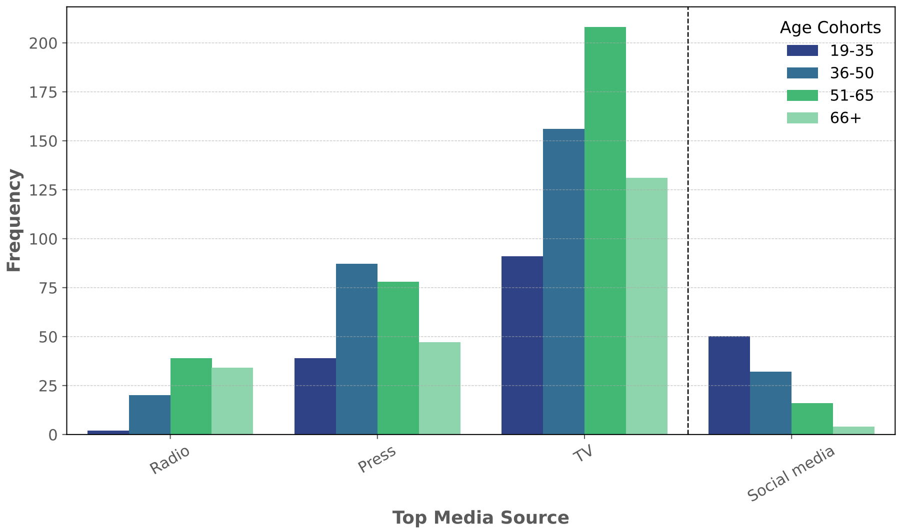
The rest of the paper is organized as follows. In section 2, I briefly summarize the Spanish political and TV landscape. Section 3 introduces our data and describes the text analysis techniques we employ, along with some descriptive statistics on both the content and audience sides. In section 5, I describe 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 I incorporate a supply side estimation.

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 Spain and age groups, as reported by the 2023 survey from CIS. Traditional media are represented in the left panel as radio, press and television. Television remains the dominant source of political information for cohorts of 50 years and onward and dominates social media for all but the youngest cohort. This pattern, however, is not specific to Spain. Figure 8 shows that television remains the dominant source of political information in Europe, particularly among older age groups. While younger cohorts (15-24, 25-39) rely more on social media, TV consistently outweighs social media among those aged 40 and above, with the gap widening significantly in the 55+ group. This pattern highlights television's enduring influence in European media landscapes, especially for older demographics, despite the rise of digital platforms.

Figure 1: Top media source to acquire political information by age cohorts



Notes: Histogram on the preferred media used for political information in Spain. Source: Barómetro CIS, 2023.

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.

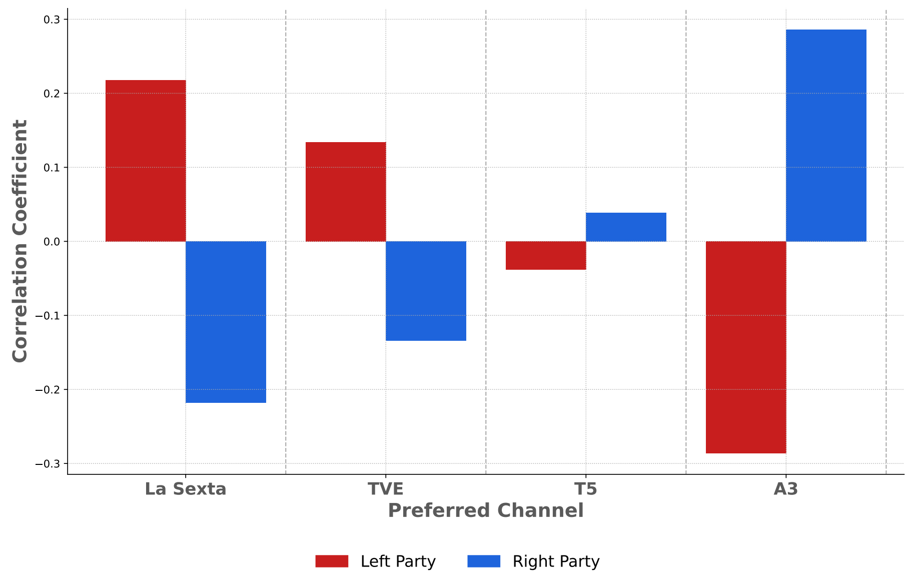
We focus only on the night edition of TV news programs for several reasons. First, it is the so-called "golden time" of the day, attracting the maximum number of viewers on TV. Second, even though channels offer other political programs, the homogeneity of TV news allows us to make very clean comparisons. 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: Correlation between preferred channel and political party



Notes: 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.Source: Built using data from CIS's Encuesta Pre-electoral 2023

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, I 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, I have daily transcripts that I match to the minute level to see what content people were exposed to. This data spans from November 2022 to the present, on a daily frequency.

Audience data

I make use of audimeter, high-frequency audience data provided by Kantar Media. For our sample period, I observe the shares of viewers for each channel on a given day and minute. Although I don't have individual data on choices, I do have geographical disaggregation for the 16 Autonomous regions in Spain (also referred to as regions hereinafter), which I will match to survey demographics.[1]

Furthermore, I 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. I also homogenize all the outlets to have the same duration and exclude instances in which regions exhibit 0 market shares.

TV Content

I 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, I focus only on text transcripts. A more detailed explanation of the entire downloading pipeline can be seen in Appendix Figure 7.

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

Agencia EFE

I obtain all news stories provided in Spanish by one of the largest news agencies in the world, Agencia EFE. Due to access limitations, I receive only the title of each story along with a short summary segment. The sample contains a total of 41K stories.

Survey Data

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

Weather

I use meteorological data of daily precipitations per Autonomous Community on the time-span matching the TV news programs (18h-00h) from the Spanish Meteorological Agency (AEMET).

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.[2] 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 [3]

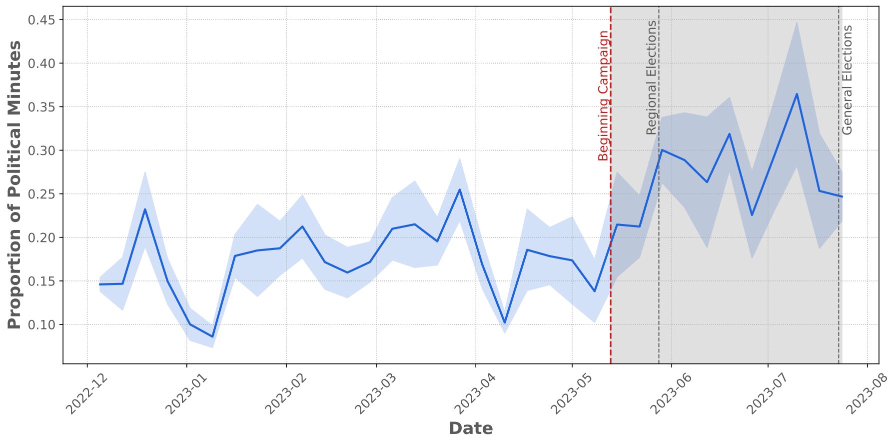
Due to sample restrictions, I divide our time span into two. The *pre-campaign* period starts at the beginning of our data collection in December 2022 and extends to the start of the first publicly announced political campaign on May 13, 2023. The *campaign* period covers both regional and general election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

[2]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, I refer to UP as either Podemos or SUMAR after its creation.

[3] I refer readers to the Spanish Media Monitor webpage to explore various metrics of coverage across political parties and actors.

Figure 3 shows the (weekly) average proportion of time devoted to national politics over time; where the *campaign* period corresponds to the shaded time span and vertical dashed lines mark the beginning of the campaign, the regional and the national elections; respectively. A channel is defined to spend one minute on national politics if the associated text for that minute contains any of the party matches in the dictionaries in table 9. Pre-campaign, channels spend on average nearly 18% of their daily time on politics. Naturally, this increases to 30% in campaign periods.

Figure 3: Proportion of time devoted to politics over time



Notes: 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.

3.2 Tone

Do channels align with their viewers' demand in the political content they offer? In this section, I describe the 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 cannot reliably discern named entities or context. For this reason, I 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, I 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, I distinguish between positive and negative tones toward the four political parties considered, and I 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 9.3.

The challenge is now to map these tone measures accounting for the intensity. I describe below how I combine air-time with tone to build content characteristics.

Building Content Characteristics

In order to create a content description of the daily programs, I consider a set up where each day a channel produces a number of stories S_{jd} indexed by s. A story is characterized by a party $p(s) \in \{L, R, \emptyset\}$, a tone $\tau(s) \in \{-1, 0, 1\}$ and the number of minutes it spans; $min(s)$. Then the proportion of positive and negative minutes together with the proportion of time devoted to national politics is defined as:

$$
\begin{aligned}
x_{jd}^{party+} &= \frac{1}{min_{jd}} \sum_{s \in S_{jd}} \Big(\mathbb{1}\{\tau(s) = 1\} \times \mathbb{1}\{p(s) = party\} \times min(s) \Big) \quad \forall party \in \{L, R\} \\
x_{jd}^{party-} &= \frac{1}{min_{jd}} \sum_{s \in S_{jd}} \Big(\mathbb{1}\{\tau(s) = -1\} \times \mathbb{1}\{p(s) = party\} \times min(s) \Big) \quad \forall party \in \{L, R\} \\
x_{jd}^{political} &= \frac{1}{min_{jd}} \sum_{s \in S_{jd}} \Big(\mathbb{1}\{p(s) \neq \emptyset\} \times min(s) \Big)
\end{aligned} \tag{1}
$$

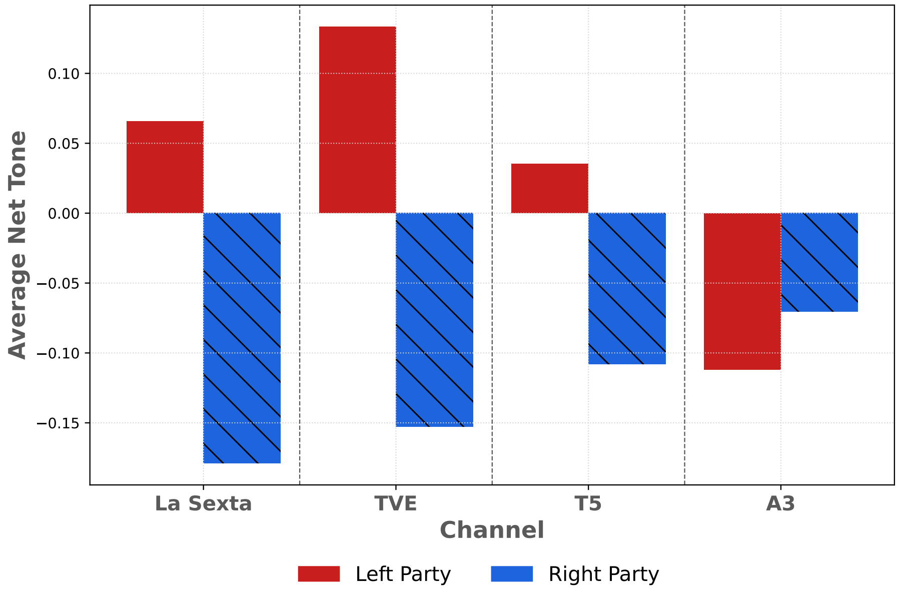
These variables will constitute the main controls on the empirical application in the next section. Figure 4 shows the net average tone relative to total time spent on politics $\frac{\bar{x}_j^{party+} - \bar{x}_j^{party+}}{\bar{x}_j^{political}}$ for each party-outlet. This measure acounts both for intensity as it takes into account the minutes that each channel spends on each category and controls for their overall time spent on poliics. Whilst all channels present a negative net tone on the right, there are still significant differences that allow to map them into a different political spectrum. Both La Sexta and the public channel, TVE, offer more pro-left content. Telecinco appears in the middle but still mantains a pro-left stance and Antena 3 is the only one with a more pro-right balance. Both Figures 2 and 4 reinforce the idea that viewers tend to self-select into channels that align with their political preferences, further illustrating the patterns of ideological segregation in media consumption.

Robustness of the text classification

Two main concerns arise with the use of LLMs as text classifiers. First, one can wonder about performance with other text classification methods. Second, LLM's have been shown to suffer from potential stochasticity making results unstable when using multiple runs of the same prompt.

To tackle the first one, I follow previous methods that used congress speeches to measure similarity to declared party labels (Gentzkow et al., 2011) (Laver et al., 2003). Similar to

Figure 4: Average sentiment across channels and parties



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

Laver et al. (2003), I build a simple word score based on the top 100 words that mostly distinguish left and right congress speeches. Figure **??** shows the normalized ideology for each of the outlets, which reinforces the left-right positions obtained by the ChatGPT classification in figure 4.

Due to their inherent stochasticity, repeated queries using the same prompt may yield different classifications. As shown in Brown and Lee (2024), this variability can introduce noise in tasks that require high consistency, particularly in content classification. To mitigate this issue, I leverage OpenAI's "functions" tool, which constrains the classifier's responses to predefined discrete, numerical outputs, reducing potential inconsistencies. Table 10 presents the mean classification scores from 100 iterations of a random sample of political stories, along with the corresponding standard errors. The relatively small standard errors suggest that despite the model's stochastic nature, the classification remains stable, in line with the findings of Doe and Smith (2024), who recommend running multiple iterations when using LLMs as classifiers.

The second concern is more fundamental as it challenges the validity of the LLM reasoning as a valid tool for political classification. To address the robustness of its results, I compare the classification results of the LLMs methodologies employed in previous works. I make use of all Spanish congress speeches during my sample period and exploit party labels to see similarities with the outlet's content (Gentzkow and Shapiro, 2010) (Laver et al., 2003). The digression in Appendix 9.4 explains the methodology in detail. Importantly, figure 14 represents the left-right mapped positions under this method which consistently map the ChatGPT classification in figure 4 and confirms the power of this approach.

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, I explore 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 9, 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.

5 Market set up

The TV market set up differs from the classical demand estimation problem as there is absence of prices and channels differentiate themselves varying product characteristics. I estimate demand using a mixed logit model (Berry, 1994) [4].

An individual i a region d chooses an outlet j to watch at the beginning of day d based on the following indirect utility :

$$U_{ijrd} = \underbrace{\sum_k x^k_{jd-1}\alpha^k + w_{rd}\gamma + \xi_{jrd}}_{\delta_{jrd}} + \underbrace{\sum_k x^k_{jd-1}\left(\sigma^k \nu^k_{ird} + \pi^k y_{ir}\right)}_{\mu_{ijrd}} + \epsilon_{ijrd} \tag{2}$$

where x^k_{jd} represents the $k \in \{R+, R-, L+, L-, political\}$ characteristic on channel j and day d as defined in 1. σ^k is a shift on characteristic k mean preferences according to unobserved preferences $\nu^k_{ird} \sim N(0,1)$. The characteristic w_t measures the precipitation level on a given day-region. I also introduce as demographics, y_{ir} that represent whether individual i is right wing or not according to survey data. Thus the parameters π^k capture

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

polarization behavior where more right-wing viewers have asymmetric tastes for content on their own and the opposed party tone.

I decompose the unobserved characteristics into $\xi_{jrd} = \xi_j + \xi_{dow} + \Delta\xi_{jrd}$, where I include product dummies that account for unobserved quality factors and day of the week dummies to control for the seasonality in our audience data.

The outside option is modeled in terms of *potential* audience. Formally, the observed share of people consuming a channel is $s_{jrd} \equiv \frac{q_{jrd}}{L_r}$where q_{jrd} is the (initial) total number of viewers watching channel j and L_r is the potential of viewers having access to TV on region r; defined by Kantar media as the population that has access to television in a given region. Importantly, I fit the model using data from the first 5 minutes of the day, where viewers have not been exposed to the daily content yet. This timing will become important for the supply identification later on.

Market shares can be approximated by a sum over the individuals in a given market, $\mathcal{I}_{rd}$ as:

$$s_{jrd} = \int d_{ijrd}(\boldsymbol{\delta_{rd}}, \boldsymbol{\mu_{ird}}) d\boldsymbol{\mu}_{ird} d\boldsymbol{\epsilon}_{ird} \tag{3}$$

where d_{ijrd} equals 1 if $U_{ijrd} > U_{ikrd} \quad \forall j \neq k$ and 0 otherwise. Survey weights are included in the integration.

5.1 Content endogeneity and News Shocks

Content characteristics are an equilibrium outcome of the differentiation game therefore creating endogeneity problems due to simultaneity bias. If, for example, local shocks to viewers' preferences are taken into account by the outlets, this would shift their content production making $\mathbb{E}(\mathbf{x}_{jrd-1}\xi_{jrd}) \neq \mathbf{0}$. Figure 10 shows the political density in the left right spectrum of each channels' audience using demographics y. Once again, the left-right position of the channels coincides with the content classification shown above and makes it likely that channels are aware of their audience political tastes thereby adapting their content to them.

To address this endogeneity concern, I instrument for each of the k endogenous product characteristics using supply shocks. I proxy the daily news landscape by using all available stories (approximately $N \approx$ 40K) from Spain provided by Agencia EFE, one of the largest communication agencies in the world. Importantly, unlike traditional media such as newspapers, television stations rely heavily on third-party news aggregators to purchase stories, images, and related content. All the channels considered in this study collaborate with Agencia EFE, thereby minimizing concerns about the agency having a specific political

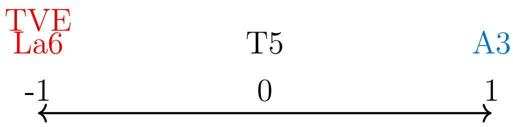


Figure 5: Illustration of the channel's equilibrium position in the political spectrum according to their average reported tone on left and right channels.

bias or being uninformed[5].

Analogous to the content covariates in 1, I control for the political composition of the news landscape using the relative number of stories available for each of the K characteristics:

$$
\begin{aligned}
z_d^{party+} &= \frac{1}{|S_d|}\sum_{s\in S_d}\Big(\mathbb{1}\{\tau(s)=1\}\times\mathbb{1}\{p(s)=party\}\Big) \qquad \forall party\in\{L,R\}\\
z_d^{party-} &= \frac{1}{|S_d|}\sum_{s\in S_d}\Big(\mathbb{1}\{\tau(s)=-1\}\times\mathbb{1}\{p(s)=party\}\Big) \qquad \forall party\in\{L,R\}\\
z_d^{political} &= \frac{1}{|S_d|}\sum_{s\in S_d}\Big(\mathbb{1}\{p(s)\neq\emptyset\}\Big)
\end{aligned} \tag{4}
$$

where I slightly abuse notation and take the relative with respect to the total number of stories in a given day produced by Agencia EFE, S_d, and tone of the stories is computed under the same ChatGPT specification and query.

Importantly for the specification, channels suffer short run shocks that impede them to place their content at their optimal long-run equilibrium. As an illustration of how this mechanism works in practice, consider the average tone results from 4. Broadly, outlets can be placed in a lef-right spectrum illustrated in figure 5. Channel would like to stick to their ideal positions but their production of stories depends on news availability. On days where the news landscape contains more stories on a particular party-tone combination, outlets are less or more constrained depending on whether this allows them to get closer to their ideal points.

Tables 7 and 8 illustrate this mechanism by displaying the stories from the most negative day for the left- and right-wing parties, respectively. The most unfavorable day for left-wing parties in the sample is February 27, 2023. On this day, a technical failure in the "ley solo sí es sí" law, promoted by the left-wing party Unidas Podemos, resulted in reduced sentences for convicted rapists. Additionally, a motion of no confidence was submitted against the Socialist Prime Minister Pedro Sánchez, and the corruption scandal known as "Caso Mediador" was affecting the government. The distribution of airtime dedicated to

[5] See collaborations for Mediaset, Atresmedia, and RTVE.

negative left-wing content varies across channels: A3 allocates nearly 19%, followed by TVE and Telecinco with 4% each, and La Sexta with only 2%.

Conversely, the day with the highest volume of negative coverage for right-wing parties was May 17, 2023. On this day, several events contributed to negative media coverage: a scandal involving the brother of Isabel Díaz Ayuso, the president of Madrid's regional government (PP), the rejection of a right-wing party's motion in the Senate, and reports of police abuse against protesters participating in a march against the extreme right party VOX. The pattern observed on February 27 is reversed in this case: La Sexta dedicates nearly 15% of its airtime to negative right-wing content, followed by TVE and Telecinco, while A3 allocates only 4%.

In order to show the previous mechanism outlined in these examples in practice, I run separate reduced form regression for each party-tone combination of the form:

$$\begin{aligned} x_{jd}^{party+} = & \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{party+}\right) \beta_j^{(party,+)} + \\ & + \sum_{party \in \mathcal{P}} \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{party-}\right) \beta_j^{(party,-)} + \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{political}\right) \delta_j + \epsilon_{jd} \end{aligned} \tag{5}$$

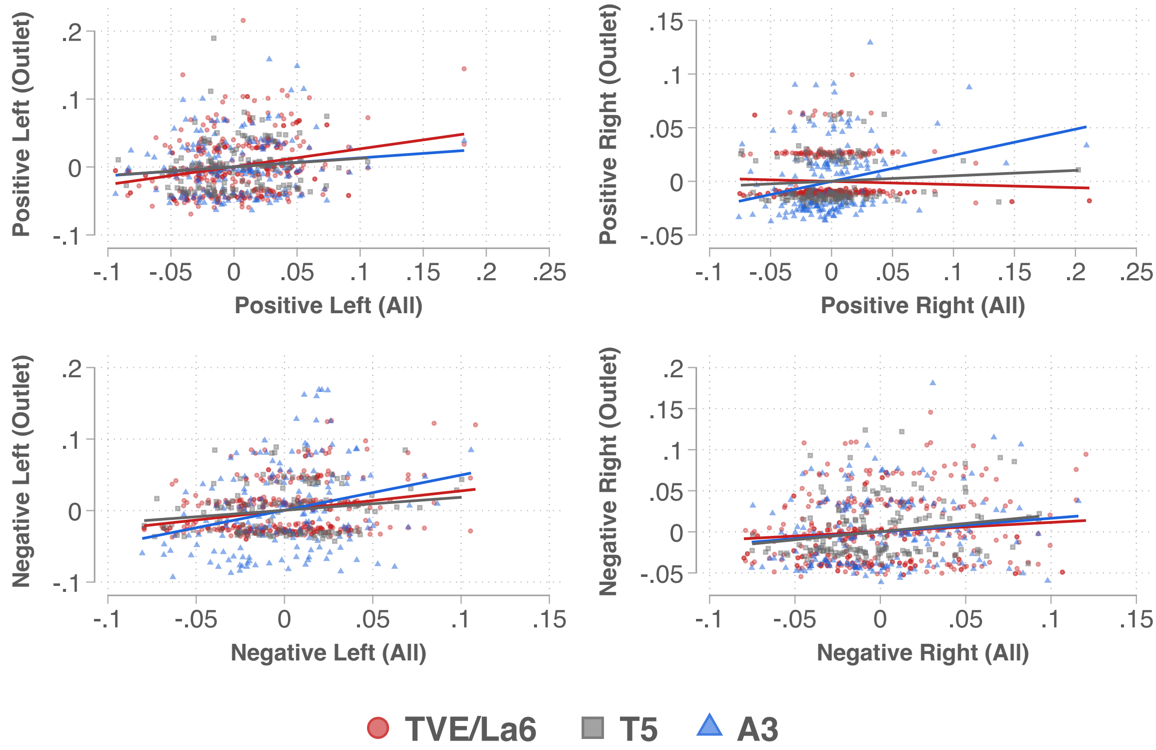
$$\begin{aligned} x_{jd}^{party-} = & \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{party+}\right) \phi_j^{(party,+)} + \\ & + \sum_{party \in \mathcal{P}} \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{party-}\right) \phi_j^{(party,-)} + \sum_{j' \in \mathcal{J}} \left(d(j')_j \times z_d^{political}\right) \rho_j + \epsilon_{jd} \end{aligned} \tag{6}$$

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

Figure 6 shows the added variable plots from equations 6 and 5 for each channel. Rows correspond to the dependent variable x_{jd}^{party+} or x_{jd}^{party-} and columns to z_d^{party+} and z_d^{party-}, respectively.

An increase of 1pp. of negative left stories is associated with a 0.49pp increase in relative time devoted to negative left wing content in the right wing channel versus a 0.27 for the left channels, ceteris paribus. The reversed situation happens when we look at positive left content. Left wing channels now produce a higher content compared to the right wing channel. These patterns alltogether illustrate the effectiveness of the shocks in practice: Outlets increase production of a particular party-tone content whenever this allows them to stay closer to their long run equilibrium.

Figure 6: Added variable plots for production of political content



Notes: The figure shows the added variable plots from the estimation of equations 6 and 5. The x axis represents $(z_d^{party,+}, z_d^{party,-})$ and the y axis the corresponding $(x_{jd}^{party,+}, x_{jd}^{party,-})$. Channels are pooled into left (TVE and La6), middle (T5) and right (A3) for visualization purposes.

Table 1: F and t-tests

Characteristic	F-test	F-test p-value	T-test	T-test p-value
$L-$	2.69	0.0686	3.51	0.0613
$L+$	1.31	0.2711	1.87	0.1715
$R-$	0.42	0.6594	0.34	0.5592
$R+$	8.78	0.0002	17.55	0.0000

Notes: Summary of F- and T-test results for equality of news variable coefficients across channels. The F-test evaluates overall equality across channels, while the T-test compares the coefficient of the main independent variable between A3 and TVE/La6.

Based on the previous discussion, I use $\left(z_d^{party+}, z_d^{party+}, z_d^{political}\right)_{party\in\{L,R\}}$ as instruments for the linear characteristics. For the K non linear coefficients that represent the variance of the individual heterogeneity I follow Gandhi and Houde (2019) and target predicted product differentiation instruments of the form $\left(\hat{x}_{jd}^k - \sum_{l\neq j} \hat{x}_{ld}^k\right)^2$, where $\hat{x}_{jd}^k$ is the predicted kth characteristic from the first stage of the instrumental variable regression 4. To target the demographic parameters, I interact the mean proportion of right votes with those instruments: $\bar{y}_r \left(\hat{x}^k_{jd} - \sum_{l\neq j} \hat{x}^k_{ld}\right)^2$.

6 Results

I present results of the BLP estimation for the pre-campaign and campaign periods in table ??. I 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 ($\sigma^{\text{R-}}$) 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 overall is negative with right wing viewers having a higher demand for it.

The average own-elasticities for each characteristic are shown in table 4. Given that increasing the positive relative minutes to a party also implies increasing the overall total minutes on politics, I compute them as:

$$\bar{\epsilon}^k = \frac{1}{J_t}\frac{1}{T}\sum_j\sum_t\left(\frac{\partial s_{jt}}{\partial x^k_{jt}} + \frac{\partial s_{jt}}{\partial x^{political}_{jt}}\right) \quad \forall k \in \{R+, R-, L+, L-\} \tag{7}$$

Variable	Pre-Campaign	Campaign
x^{L+}	-0.046	-0.299
x^{L-}	0.050	0.219
x^{R+}	0.034	-0.006
x^{R-}	-0.033	0.045

Table 4: The table shows the estimated elasticities for the political characteristics from equation 7.

7 Supply Side: Work in progress

In this section, I model the channels' editorial decisions. The following quote by A3 news director Vicente Vallés clearly illustrates their slant decision process. Given that the audience information available to outlets is similar to the information used in our demand estimation, I model their production decisions as a pure horizontal differentiation game. In this framework, channels choose the content to offer on a given day while incurring searching costs for producing stories of a given type. Intuitively, the goal of the outlets is to retain as much audience as possible with the content they offer.

> "Even in a news program, it's deeply subject to the daily variation of audiences. It's measured minute by minute how the interest of the viewers shifts and how a drop of half a point when tackling a current issue can provoke the decision not to revisit that topic... simply because audience is lost, and thus, profitability."
>
> – Vicente Vallés, Director and presenter of Antena 3 TV News, 2014[6]

Key for identification is the timing of preferences as outlined in the demand estimation. Audience preference for which channel to start watching the news is driven by the content presented on the previous day, since viewers have not yet been exposed to the current day's content. Since outlets must decide which content to offer on a given day based solely on audience information from the previous day, there is no simultaneity in their editorial production and **[further explanation needed]**.

The maximization problem for a given channel j on day d is:

$$\begin{aligned} &\max_{\{\mathbf{x}_{jd}\}} \left\{ \sum_r s_{jrd+1}(\boldsymbol{x}_{jd}, \boldsymbol{x}_{-jd}) \frac{L_r}{L} - \sum_k \lambda^k \mathcal{C}(x^k_{jd}, z^k_d) + \boldsymbol{\eta}_{jd} \boldsymbol{x}_{jd} \right\} \\ &s.t. \quad x^{L+}_{jd} + x^{R+}_{jd} + x^{L-}_{jd} + x^{R-}_{jd} + x^{\emptyset}_{jd} = x^{political}_{jd} \\ &\qquad\; x^k_{jd} \in [0,1] \quad \forall k \end{aligned} \tag{8}$$

Channels want to retain maximum amount of audience over all the regions :$\sum_r s_{jrd+1}(\boldsymbol{x}_{jd}, \boldsymbol{x}_{-jd}) \frac{L_r}{L}$ who has preferences over the content they offered on that day: $\mathbf{x}_{jd}$.

Outlets face searching cost of producing stories analogous to those suggested in (Simonov and Rao, 2022). The more stories of a type there are available on a given day, i.e high z^k_d , the cheaper it is for them to produce them. I model a simple functional form $\mathcal{C}(x^k_{jd}, z^k_d) = \frac{(x^k_{jd})^2}{z^k_d}$ where the inverse of the available stories acts analogously to input prices. Te goal is to estimate each of the parameters λ^k that controls for the asymmetric production costs of each type of content.

[6] Source: https://cadenaser.com/ser/2014/12/03/television/1417630810_539829.html

Rearranging the first order conditions [7] and normalizing by the neutral category $k = \emptyset$ I get:

$$\frac{\partial \mathcal{C}(\cdot)}{\partial x_{jd}^{k}} = \frac{\lambda^{\emptyset}}{\lambda^{k}} \sum_{r} \frac{\partial s_{jrd+1}}{\partial x_{jd}^{k}} L_r + \xi_{jd}^{k} \quad \forall j, k \tag{9}$$

Equation 9 can be estimated with GMM; and it is a particular case of the *common coefficients* GMM estimation in Hayashi (2000).

7.1 Cost estimates

8 Conclusion

Understanding the demand of political information is crucial to understand political polarization and media market regulation. However, endogeneity concerns often impede classical demand estimation techniques due to the lack of valid instruments. In this paper, I introduce a novel dataset that comprises the Spanish TV market, where I match daily transcripts of the TV news to audimeter data on viewership. I propose a new methodology that makes uses of text analysis and Large Language Models (LLMS) to analyze the production of political content in TV news. This methodology extracts the political tone and intensity of the daily TV news and exploits the random availability of political events, together with channel's long-run ideological positions, to measure supply shocks that allow the estimation of demand preferences.

I show that channels face asymmetric constraints in their production of political content depending on whether the composition of the day is more or less favorable to their ideological stance. I estimate a structural BLP demand model where I split my sample into both pre-campaign and during political campaign periods. This model allows me to introduce heterogeneity into the demand estimation and decompose political preferences based on the ideological composition of the audience. My results reveal that while there is no significant asymmetric demand for political content during the pre-campaign period, affective polarization emerges during the political campaign, with right-wing viewers demanding more negative content about the opposing party and more favorable content about their own.

[7] Notice that constrains can be ignored for estimation but should be recovered when the game is use for counterfactual analysis.

Coefficient	Parameter	Estimate	Std. Error
	Pre-campaign		
Positive Left	β^{L+}	-14.19	(23.16)
Positive Right	β^{R+}	-21.17	(45.04)
Negative Left	β^{L-}	45.26**	(19.82)
Negative Right	β^{R-}	41.82*	(23.91)
Political	$\beta^{political}$	8.23	(6.62)
Weather	γ	-0.01	(0.02)
Positive Left	σ^{L+}	0.82	(397.80)
Positive Right	σ^{R+}	0.77	(453.14)
Negative Left	σ^{L-}	29.22***	(6.01)
Negative Right	σ^{R-}	23.38***	(4.39)
Political	$\sigma^{political}$	0.00	(70.70)
Right-Wing × Positive Left	π^{L+}	25.79	(33.64)
Right-Wing × Positive Right	π^{R+}	91.36	(117.38)
Right-Wing × Negative Left	π^{L-}	-69.55**	(32.71)
Right-Wing × Negative Right	π^{R-}	-87.95	(67.72)
Right-Wing × Political	$\pi^{political}$	-21.63	(17.22)
	Campaign		
Positive Left	β^{L+}	144.87*	(78.34)
Positive Right	β^{R+}	-122.19**	(51.20)
Negative Left	β^{L-}	-105.33**	(45.34)
Negative Right	β^{R-}	99.64**	(50.52)
Political	$\beta^{political}$	-6.93*	(3.82)
Weather	γ	0.00	(0.01)
Positive Left	σ^{L+}	0.00	(327.82)
Positive Right	σ^{R+}	10.57	(14.31)
Negative Left	σ^{L-}	0.20	(120.18)
Negative Right	σ^{R-}	0.08	(57.87)
Political	$\sigma^{political}$	0.00	(21.75)
Right-Wing × Positive Left	π^{L+}	-458.00**	(215.29)
Right-Wing × Positive Right	π^{R+}	335.88**	(140.29)
Right-Wing × Negative Left	π^{L-}	293.56**	(131.61)
Right-Wing × Negative Right	π^{R-}	-300.91**	(129.69)
Right-Wing × Political	$\pi^{political}$	20.55*	(10.56)

Table 2: BLP Estimation Results with Standard Errors

The table shows the results of the BLP estimation of model 2. The estimations are divided into 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. The total number of observations are $N_{campaign} = 2307$ and $N_{pre\_campaign} = 6604$.

Coefficient	Parameter	Estimate	Std. Error
	Pre-campaign		
Positive Left	β^{L+}	-14.23	(66.23)
Positive Right	β^{R+}	-22.22	(104.36)
Negative Left	β^{L-}	-46.87*	(26.03)
Negative Right	β^{R-}	-48.78*	(26.15)
Political	$\beta^{political}$	8.93	(12.42)
Weather	γ	-0.01	(0.04)
Positive Left	σ^{L+}	0.19	(1402.79)
Positive Right	σ^{R+}	0.90	(517.58)
Negative Left	σ^{L-}	28.98	(24.18)
Negative Right	σ^{R-}	25.43***	(8.34)
Political	$\sigma^{political}$	0.06	(8.13)
Right-Wing × Positive Left	π^{L+}	26.02	(127.62)
Right-Wing × Positive Right	π^{R+}	95.32	(366.32)
Right-Wing × Negative Left	π^{L-}	73.17***	(28.15)
Right-Wing × Negative Right	π^{R-}	100.98	(67.13)
Right-Wing × Political	$\pi^{political}$	-23.49	(23.16)
	Campaign		
Positive Left	β^{L+}	167.41***	(61.46)
Positive Right	β^{R+}	-147.17**	(61.22)
Negative Left	β^{L-}	120.77***	(29.52)
Negative Right	β^{R-}	-114.01***	(27.21)
Political	$\beta^{political}$	-9.67**	(4.48)
Weather	γ	0.01	(0.02)
Positive Left	σ^{L+}	0.06	(355.44)
Positive Right	σ^{R+}	21.05	(17.47)
Negative Left	σ^{L-}	0.00	(394.60)
Negative Right	σ^{R-}	0.01	(182.36)
Political	$\sigma^{political}$	2.37	(4.57)
Right-Wing × Positive Left	π^{L+}	-533.26***	(142.68)
Right-Wing × Positive Right	π^{R+}	396.27***	(116.17)
Right-Wing × Negative Left	π^{L-}	-340.36***	(91.12)
Right-Wing × Negative Right	π^{R-}	345.05***	(84.14)
Right-Wing × Political	$\pi^{political}$	24.06***	(6.13)

Table 3: BLP Estimation Results with Standard Errors

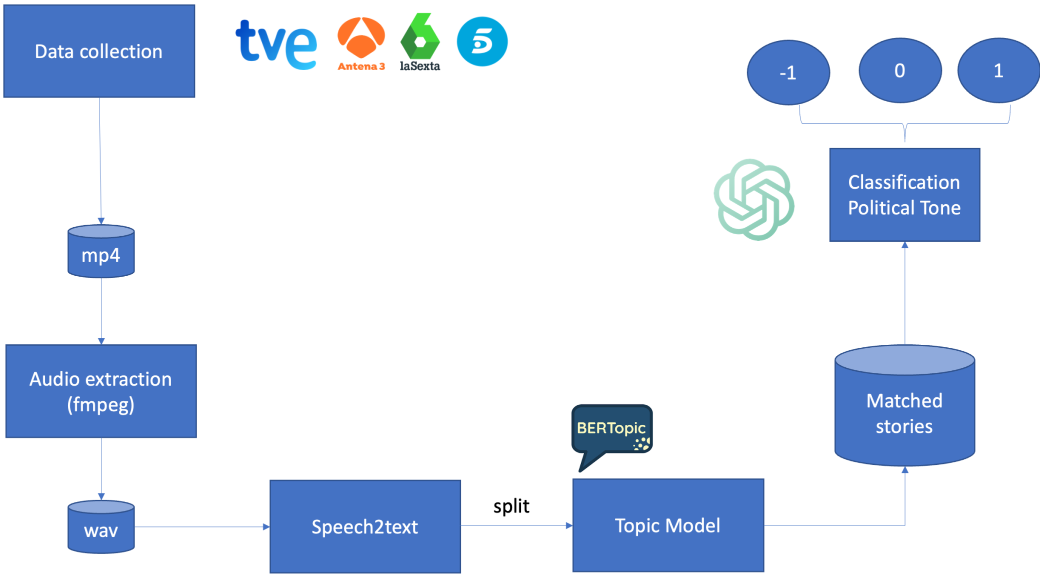
The table shows the results of the BLP estimation of model 2. The estimations are divided into the pre-campaign and campaign period. Both day-of-the-week and outlet fixed effects are included. Standard errors are clustered at the region level. The total number of observations are $N_{campaign} = 2307$ and $N_{pre\_campaign} = 6604$.

9 Appendix

9.1 Figures

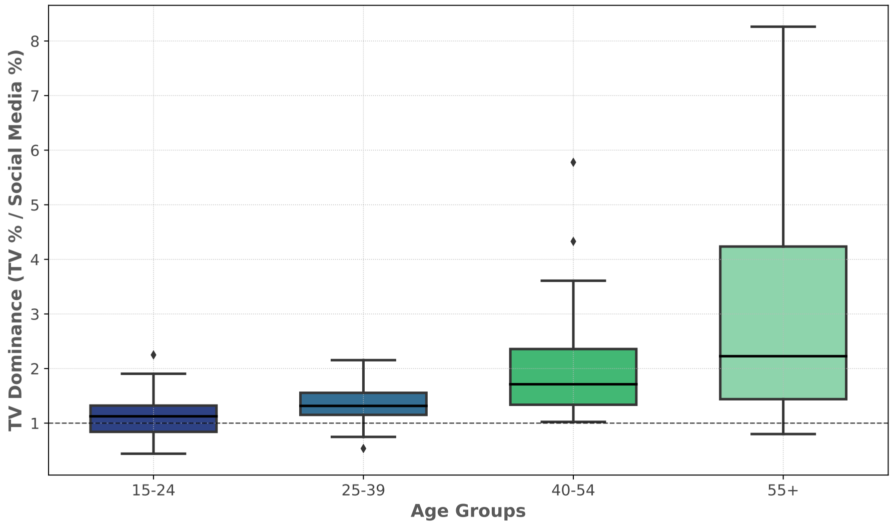
Pipeline:

Figure 7: Pipeline for content downloading and classification



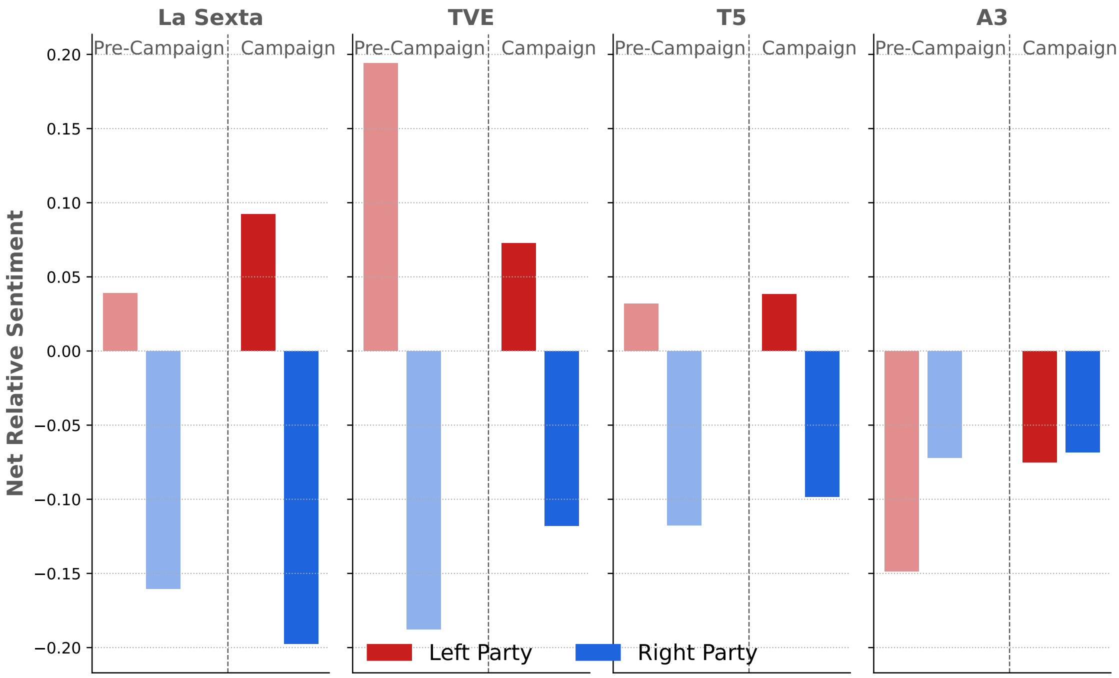
Notes: Pipeline for the text downloading. First videos are downloaded daily from the main TV channels. Google engine is used to convert the audio to text. I then split the stories by minute and user BERTopic to classify and match them. Finally, ChatGPT4 is used to classify political tone.

Figure 8: Box plot for TV dominance across countries and age cohorts



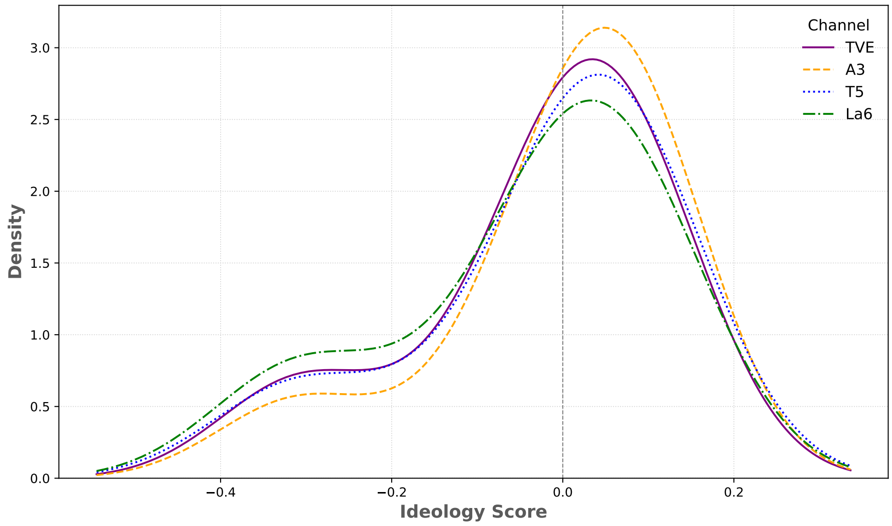
Notes: Box plot of the relative TV dominance relative to social media by age cohorts. Using data for the 27 countries in Eurobarometer with $N = 112059$. Relative dominance is calculated as the total number of respondents that mark TV as the main source for political information relative to those that mark social media. Source: Eurobarometer, 2022.

Figure 9: Net sentiment across channels and parties pre and during campaign



Notes: The figure shows the relative tone calculated as the average sentiment over right and left parties. The vertical dashed line delimits results pre and during campaign periods, respectively.

Figure 10: Density estimation for channels ideology based on audience share data



Notes: 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.

9.2 Tables

9.3 Chat GPT ideology classification

In this section I 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, I first filter our split stories using a simple dictionary based approach into those that might contain any relevant political information. Table 9 shows the key terms used to filter the political stories. After the match, I obtain a final number of 15406 political stories that I 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 9: 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.

9.4 Digression: LLMs Stability and Alternative Ideology classifications

Stability of the classifier

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 me from doing the whole approach multiple times but I show results on stability below on a random subset of 40 stories. Table 10 shows the mean and standard dviation of the classification output for 100 rounds of classification.

Results of the final classification for the non neutral stories are shown in figure 11. Each bar represents the percentage of stories of that given sentiment associated to each political

Table 5: Top 5 Topics by party-tone category on Agencia EFE after BERTopic.

Topic Words	Count
Positive Right	
vox, motion, abascal, censure, pp, tamames, santiago, garriga, parties, party	312
feijóo, núñez, alberto, pp, leader, gamarra, parties, party, general, cuca	113
guardiola, extremadura, mérida, maría, vara, extremeño, council, assembly, candidate	68
mazón, valencian, president, corts, valencian, valencians, government, carlos	64
electoral, elections, general, jec, board, campaign, 28m, 23j, vote	63
Negative Right	
vox, motion, abascal, censure, pp, tamames, santiago, garriga, parties, party	369
abortion, castilla, healthcare, anti-abortion,pregnancy, abortions, law, mañueco	112
code, criminal, sedition, embezzlement, reform, crime, penalties, amendments	81
electoral, elections, general, jec, board, campaign, 28m, 23j, vote	71
sánchez, pedro, feijóo, president, government, núñez, leader, alberto, pp, executive	69
Positive Left	
yes, only, reform, sexual, law, is, violence, reform, just, podemos	200
sánchez, pedro, feijóo, president, government, núñez, leader, alberto, pp, executive	183
yolanda, díaz, vice president, sumar, second, podemos, labor, leader, minister	179
sumar, podemos, errejón, iu, íñigo, yolanda, parties, coalition, díaz, left-wing	142
psoe, socialists, federal, secretary, parties, socialist, lobato, espadas, congress	129
Negative Left	
yes, only, reform, sexual, law, is, violence, reform, just, podemos	182
sánchez, pedro, feijóo, president, government, núñez, leader, alberto, pp, executive	153
vox, motion, abascal, censure, pp, tamames, santiago, garriga, parties, party	129
sexual, assault, sexually, sexual, minor, violence, assault, convicted, abuse	108
psoe, socialists, federal, secretary, parties, socialist, lobato, espadas, congress	72

Table 6: Table shows the top 5 topics associated with the top stories that provided more positive/negative coverage for each party in the Agencia EFE corpus. words were translated to English with the help of ChatGPT.

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.

Table 10: Mean and Standard Error for 100 Rounds of ChatGPT Classification

Statistic	PP	PSOE	VOX	UP
Mean	-0.014	0.106	-0.053	0.024
Standard Error	0.003	0.004	0.001	0.002

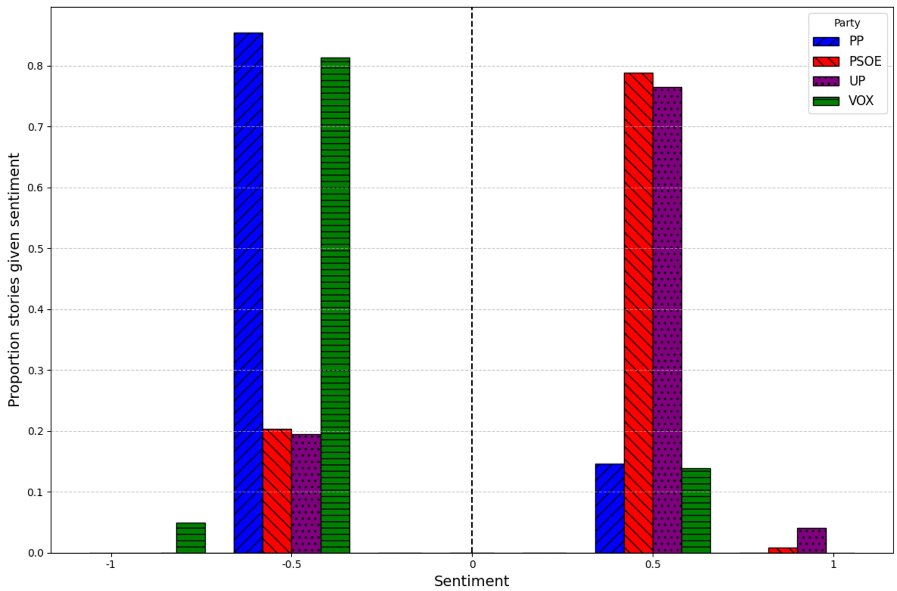
Note: The table shows the mean and standard error for 100 rounds of ChatGPT classification of political content with 40 random political stories.

Top stories for Negative Left
-Reduction of a convicted rapist's sentence in Salamanca under the "Solo sí es sí" law
-Seville Court reduces a murder and sexual assault sentence by 5 years due to the "Solo sí es sí" law
-Vox formally submits a motion of no confidence against Prime Minister Pedro Sánchez
-Madrid's regional president, Isabel Díaz Ayuso, predicts that the "Mediator Case" will bring down the government

Channel	Proportion of Negative Left
TVE	0.04
Antena 3	0.19
Telecinco	0.04
La Sexta	0.02

Table 7: Top table shows the main stories contributing to negative left content on 2023-02-27, the highest day with negative left content, summarized and translated to English by ChatGPT. Bottom table shows the proportion of minutes devoted to negative left content per channel on the same date.

Figure 11: ChatGPT distribution of scores for each political party



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

Alternative ideology classification

Top stories for Negative Right
-Congress declarations against Ayuso over alleged "bribes" to her brother
-Marinaleda criticizes the "abusive" arrest of two residents during a protest against Vox
-The Senate rejects a PP motion on the government's alleged partisan use of the Falcon jet

Channel	Proportion of Negative Right
TVE	0.074
Antena 3	0.038
Telecinco	0.077
La Sexta	0.148

Table 8: Top table shows the main stories contributing to negative right content on 2023-05-17, the highest day with negative right content, summarized and translated to English by ChatGP. Bottom table shows the proportion of minutes devoted to negative right content per channel on the same date.

Similar to Gentzkow and Shapiro (2010) and Laver et al. (2003) I exploit party ideology in congress speeches to calculate similarity measures of the outlet's speech. I make use of all congressional speeches produced during my sample period and associate each speaker with their respective political party, filtering to retain the set of relevant parties.

I follow a similar, non-linear version of Laver et al. (2003) and create a score for each word w in the congress speech as:

$$\text{Score}(w) = \ln\left(\frac{\text{freq}(w, \text{Left}) + \alpha}{\text{Total}_{\text{Left}} + \alpha}\right) - \ln\left(\frac{\text{freq}(w, \text{Right}) + \alpha}{\text{Total}_{\text{Right}} + \alpha}\right), \tag{10}$$

where:

- $\text{freq}(w, \text{Left})$ is the number of times word w appears in the speeches of left-leaning parties (PSOE and UP),
- $\text{Total}_{\text{Left}}$ is the total word count in left-party speeches,
- $\text{freq}(w, \text{Right})$ and $\text{Total}_{\text{Right}}$ are defined analogously for right-leaning parties (PP and Vox), and
- α is a small smoothing parameter.

I select the value of alpha that maximizes accuracy of label prediction in the congress dataset; $\alpha = 0.9$. Words with high positive scores are used disproportionately in left-leaning speeches, while those with high negative scores are more characteristic of right-leaning speeches. I rank all words by their computed score and select the top 100 left-coded words and top 100 right-coded words, represented in wordcloud figures ?? and ??.

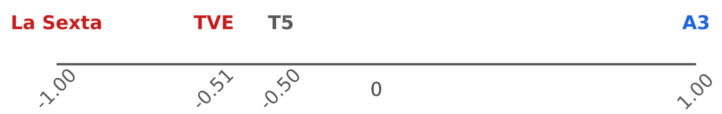
Figure 12: Word Cloud Top Left Words

Figure 13: Word Cloud Top Right Words

Notes: The wordclouds represent the Top Words with lowers (left) and highest (right) scores as defined in equation 10. Size of the word is weighted by word frequency appearance.

Finally, to classify the channel data, I calculate for each channel the fraction of tokens that appear in the left-coded list versus the right-coded list. This yields an index of ideological slant that reflects which side's language is more prevalent in the channel's content. The

Figure 14: Normalized Ideology Scores



Notes: The figure shows the normalized ideology positions of the channels after the text classification based on congress speeches.

9.5 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, I propose a simple test of this behavior in a 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, I make sure that channel's content has not play a role yet. Specifically, I run channel by channel regressions of the form:

$$ln\left(\frac{q_{jt}}{L_r - Q_{jt}}\right) = \beta_0 + \sum_{p\in\{L,R\}} \left(\rho^{p+} z_d^{p+} + \rho^{p-} z_d^{p-}\right) + \phi z_d^{political} + \gamma_{dow} + \gamma_r + \epsilon_{jt} \tag{11}$$

	$ln\left(\frac{q_{jrd}}{L_r - Q_{rd}}\right)$			
	(TVE)	(A3)	(T5)	(La Sexta)
ρ^{R+}	-1.1830***	-1.2970***	-0.9093***	-0.9572***
	(0.2379)	(0.2059)	(0.2365)	(0.2832)
ρ^{L+}	0.2434	0.8154***	0.6709***	0.6657***
	(0.2172)	(0.1709)	(0.2461)	(0.2339)
ρ^{R-}	-0.6172***	-0.7594***	-1.0197***	-0.5473**
	(0.2177)	(0.1771)	(0.2712)	(0.2548)
ρ^{L-}	1.5192***	1.7982***	1.0508***	1.0127***
	(0.2582)	(0.2026)	(0.2770)	(0.2861)
$\rho^{political}$	-0.5071**	-0.5051***	0.0565	-0.3992*
	(0.2029)	(0.1723)	(0.2213)	(0.2388)
Constant	-3.3131***	-2.7860***	-3.6418***	-4.1093***
	(0.0650)	(0.0546)	(0.0671)	(0.0724)
N	2231	2555	2001	2381

Robust standard errors in parentheses

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

Table 11: The table shows the results from the estimation of 11 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.

Table 11 show the estimated coefficients from equation 11 where each column corresponds to a separate outlet regression and I condition on the minute zero audience. Equation 11 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.

9.6 Alternative methods for segment splitting in unstructured text

I explain here different unsupervised text splitting methods that were tried to split the segments of the day. To the best of my 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 me to use them for the 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, I 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 15), which is the area where the output is expected to appear.

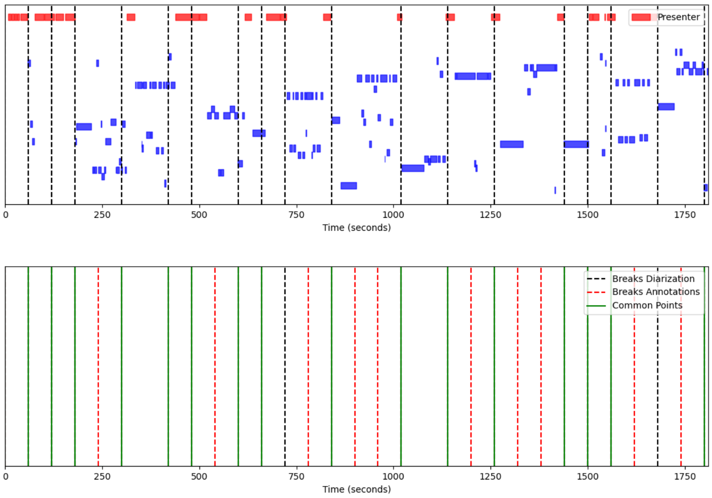
Figure 15: Example of image story delimiter

Notes: 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*, I 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 16 illustrates this procedure and the comparison with the manually annotated labels for an example day-channel.

Figure 16: Comparison audio splitting with annotated section splits



Notes: 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.

Brown, E. and Lee, D. (2024). Decoding unstructured text: Enhancing llm classification accuracy with redundancy and confidence. *Conference on AI Applications in NLP*.

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

Doe, J. and Smith, A. (2024). Llm stability: A detailed analysis with some surprises. *arXiv preprint arXiv:2408.04667*.

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. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

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.

Hayashi, F. (2000). *Econometrics*. Princeton University Press, Princeton, NJ. See Chapter 8, p. 286 for common-coefficients GMM discussion.

Laver, M., Benoit, K., and Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2):311–331.

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.

Simonov, A. and Rao, J. (2022). Demand for online news under government control: Evidence from russia. *Journal of Political Economy*, 130(2):259–309.

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*.