

# The Impact of Political Campaigns on Demand for Partisan News

Luis Menéndez \*

May 14, 2025

## Abstract

Political polarization in news consumption has recently gained attention, yet policies to limit it are hard to evaluate. This paper introduces a novel, self-collected dataset on Spanish prime-time TV news—the dominant source of political information in Europe. I identify stories minute-by-minute and match them across newscasts to compare editorial treatment. Combining machine-learning methods with large language models, I classify the partisan slant of each story and use the resulting data to document changes in news coverage during the 2023 presidential campaign. I then match these data to high-frequency audience-meter records and estimate a structural random-coefficients demand model to understand how outlets compete in coverage and tone of news about different political parties. Exogenous shifts in the news landscape that constrain channels’ ability to shape content serve as instruments for slant and allow me to recover demand preferences. I find significant evidence of affective polarization only after the election campaign begins. Given the demand estimates, I back out outlets’ content preferences from a horizontal-differentiation game. Finally, I run counterfactual simulations to assess the effects of adopting the proportional-airtime rules enforced in other European countries but not in Spain. The framework offers a tractable tool for evaluating media policies aimed at content fairness or fighting misinformation.

---

\*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 (Zhuravskaya et al., 2020). Consequently, policymakers often regulate media markets to ensure informational plurality during political campaigns.<sup>1</sup>

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. 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 estimate the extent of confirmation bias and polarization in the demand for political content and how this changes with political campaigns. To this purpose I make use of a unique dataset that comprises both the demand and the supply of political content in Spanish TV news. On the demand side, I use high frequency audimeter data; thereby addressing some of the well-known limitations of survey-based studies (Prior, 2009). For the supply of content, I built a unique text dataset and use Natural Language processing and LLMs to classify the political tone of the news. This classification is robust to previous methods used in the literature but allows me to do a more flexible specification where I can distinguish the airtime devoted to each party-tone on a given day.

By exploiting geographical variation in the audience demand, I allow the preferences for content to differ by demographics- i.e political ideology. This allows me to study polarization in the news consumption.

I document the market equilibrium and the endogeneity that arises from it. More right wing outlets also match with more right wing viewers and viceversa. This makes the content offered by a channel an equilibrium outcome of a differentiation game and poses endogeneity concerns in the demand estimation. I propose a novel identification strategy that allows me to track supply shocks in the content offered by the outlets. Specifically, I obtain all the stories that have been produced in Spain by the largest Spanish news agency which supplies the content to all the outlets in my sample. I see this as a proxy for the "news landscape" available for the outlets to produce. The mechanism is then intuitive: Outlets are in a long run equilibrium in the differentiation game and want to stick to their ideological positions. Daily news shocks limit what they can produce and thus make some

---

<sup>1</sup>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).

---

days more or less favorable for a given outlet depending on the nature of the stories available. My descriptive evidence confirms this intuition: Right (left) wing channels increase their production of stories more whenever the pool of favorable right wing stories is higher.

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. XXX

In the last part of the paper I model and estimate the news production of the outlets making use of the demand estimates. Outlets play a pure horizontal differentiation game (no prices) and choose the slant of the day to maximize their viewership. Costs come from producing content as a function of the inputs they have available on that given day. Key for identification is the timing of the game. By using the initial audience of the day (i.e before viewer's have been exposed to the actual content); I avoid simultaneity issues as audience reactions follow content chosen since the later can't be modified life.

I estimate the cost parameters and solve the new equilibrium under different counterfactual scenarios evaluating the impact of recently debated regulations such as the proportional airtime rule.

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.

## Related Literature

This paper contributes to several strands of literature on media economics and political economics. The first and more close is the estimation of the demand for political information and polarization both in social and traditional media. Previous works have estimated the demand for political content on online browsing behavior (Simonov and Rao, 2022) and, more descriptively, attention has been paid to the study of consumers news outlets (Peterson, 2018) (Gambaro et al., 2021). I contribute to this literature by first making use of detailed content data that allows me to track daily variations in the stance of the channels. I also propose a novel identification strategy that relies on *external* news aggregators providers. This methodological contribution allows to find proper instruments for a broad variety of

---

media production set ups where news sources rely on external providers to filter and publish their news.

This work could also inform theoretical works on the sources of media bias. Even profit maximizing media might choose biased reports when there is a high demand for confirmatory reports. Although it goes far beyond the scope of this work, the *delegation* mechanism; where viewers need to rely on media sources that filter all the information available. In this case, perfectly rational individuals might demand confirmatory reports. My work contributes to this literature by proposing nice mechanisms to measure and estimate news production from a pool of stories; thereby tracking different source of biases (i.e filtering vs outright distortion) (Gentzkow et al., 2014).

Social media has been shown to drive polarization attitudes through the creating on online echo chambers (Bail et al. (2018) ; Zhuravskaya et al. (2020) ); I join the recent work on traditional media (Schneider-Strawczynski and Valette, 2025) and show that TV remains to be the primary source of political information in Europe and how it might contribute to polarization in key times such as political campaigns.

My paper is methodologically related to the estimation of complete market structures for political estimation. Although in a different context, Longuet-Marx (2025) also models a BLP (Berry, 1994) style demand for political candidates position that then allows him to track the candidates positioning game. I contribute to this literature by introducing a set up that has not been extensively studied before but it's at the core of social and traditional media: pure horizontal games under the absence of prices. The introduction of endogenous content impedes the treatment of product characteristics as exogenous and requires the need of high dimensional instruments that serve as supply shocks. Furthermore, thanks to the availability of high frequency data; I exploit timing on the news' decision of the channels dealing with simultaneity concerns in the cost estimation.

## Rest of the paper

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 7 shows the current work in progress where I incorporate a supply side estimation.

---

## 2 Context

Before detailing the data used and the text classification methodology I provide a brief context on the Spanish TV market and the political landscape heading to the 2023 general elections.

### Spanish TV market

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. 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.

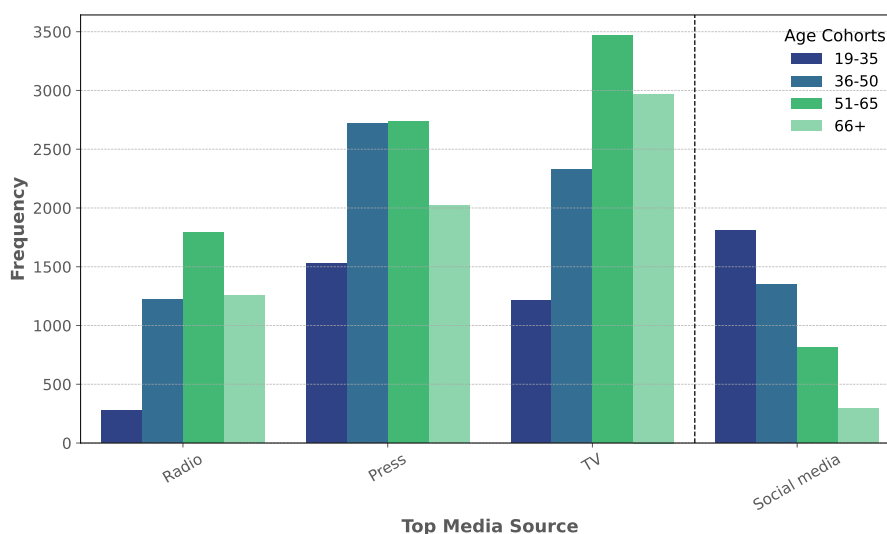
### Traditional Media and Political Information

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 10 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.

### Political Context

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

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



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

the party began to attract a substantial portion of the electorate.<sup>2</sup> 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<sup>3</sup>

### 3 Data and Descriptive Statistics

I built 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 build a scraping pipeline that records live the daily news programs and process this information onto text. This data spans from December 2022 to July 2023 on a daily frequency.

#### Audience data

I make use of audimeter, high-frequency audience data provided by Kantar Media. I observe

---

<sup>2</sup>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.

<sup>3</sup> I refer readers to the [Spanish Media Monitor](#) webpage to explore various metrics of coverage across political parties and actors.

---

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.<sup>4</sup>

## 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 9.

I 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 me to make very clean comparisons. These programs are broadcasted 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. Third, all TV news are highly fact checked and they remain the most trusted source of information in Spain<sup>5</sup>

## 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 for my sample period.

## Survey Data

To understand polarization behavior, I use survey data gathered from the Centro de Investigaciones Sociológicas (CIS). Specifically, I rely on the "intention to vote" question and map it into my binary left-right spectrum.

## Weather

---

<sup>4</sup>The Canary Islands and La Rioja are excluded due to different time zones and zero market shares; respectively.

<sup>5</sup>The Digital News Report 2023 reveals notable patterns in public trust towards news media across Europe, with Spain reflecting a stable but modest level of confidence. According to the Digital News Report 2023, the public service broadcaster RTVE and Antena 3 Noticias continue to rank as the most trusted news sources, with trust levels of 53% and 52% respectively. laSexta registers a trust increase of 2.7 percentage points over the previous year, reaching approximately 47%, and Telecinco follows at 39% (Newman et al., 2023).

---

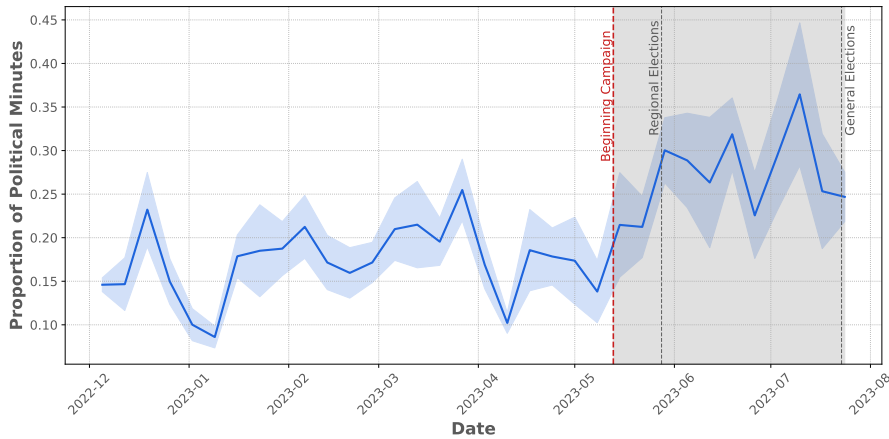
I use meteorological data of daily precipitations per Autonomous Community on the time-span matching the TV news programs (18h-00h) from the Spanish Meteorological Agency (AEMET).

### 3.1 Political Coverage and Elections

Due to sample restrictions, I divide the 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.

Figure 2 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 12. Pre-campaign, channels spend on average nearly 18% of their daily time on politics. Naturally, this increases to 30% in campaign periods.

**Figure 2:** 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 Political Tone

Here I describe the methodology used to classify the political tone of the content broadcasted. I first document the political orientation of the audiences and then use this as intuitive



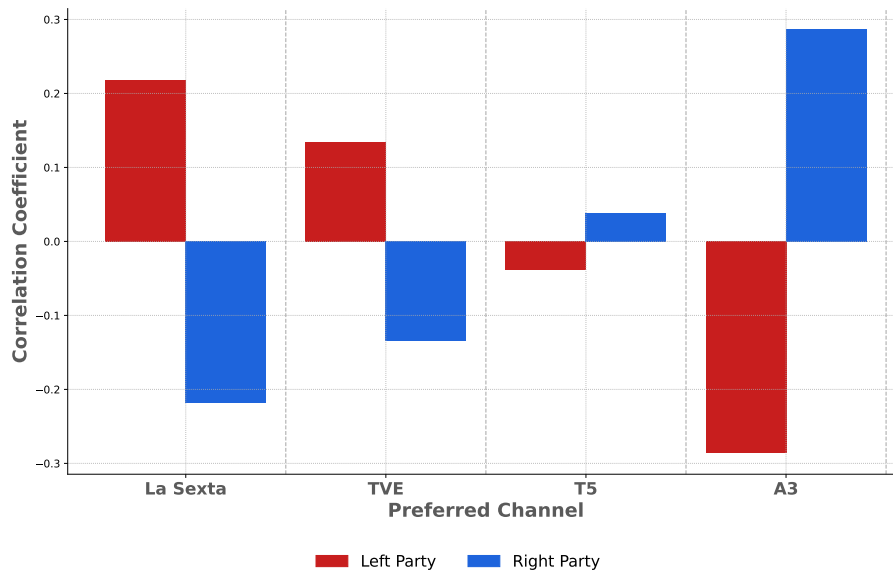
---

benchmark to validate the content after classification.

## Political orientation of the audience

Figure 3 shows the correlation between political orientation and preferred channel for acquiring political information, based on survey data from the Centro de Investigaciones Sociológicas (CIS). 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 3:** Correlation between preferred channel and political party



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

I also build a density estimate of the audience preferences based on the audience shares. Figure 12 shows the political density in the left right spectrum of each channels' audience using demographics  $y$ . The left-right position is consistent with the survey correlations shown above and allows me to use these broad preferences as a benchmark from what to expect from the political slant of the outlets.

---

## Text Classification

Do channels align with their viewers' demand in the political content they offer? 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 10.4.

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

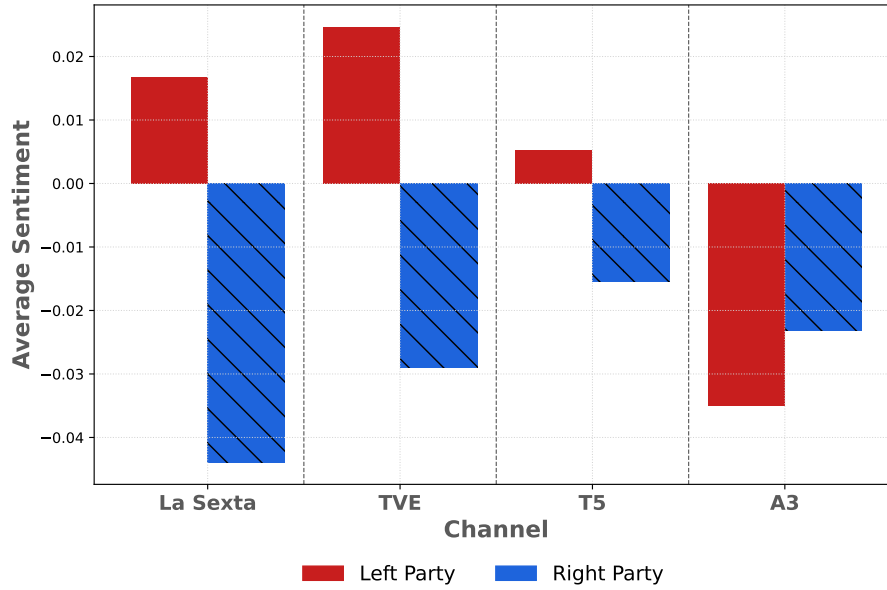
In order to create a content description of the daily programs, I consider a set up where each day a channel produces a number of stories  $S_{jd}$  indexed by  $s$ . A story is characterized by a party  $p(s) \in \{L, R, \emptyset\}$ , a tone  $\tau(s) \in \{-1, 0, 1\}$  and the number of minutes it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

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

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 accounts 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 politics. 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 maintains a pro-left stance and Antena 3 is the only one with a more pro-right balance. Both Figures 3 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.

**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.

The results match closely those from declared survey data. Figure 5 shows the normalized left-right position of the channels based on the survey data shown in 11 and the ChatGPT classification.

**Figure 5:** Normalized Ideology Scores by Channel



Notes: The figure compares normalized left-right audience positions for Spanish television channels. Panel (a) uses ChatGPT-based text classification; panel (b) uses viewer ideology data from CIS surveys. Channels are mapped into the  $[-1, 1]$  scale using their correlation (tone) and normalizing values into this scale.

---

## Robustness of the text classification

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

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 13 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 [Atil et al. \(2024\)](#), 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 10.5 explains the methodology in detail. Importantly, figure 16 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 11, 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)<sup>6</sup>.

An individual  $i$  a region  $r$  chooses an outlet  $j \in \mathcal{J} \equiv \{La6, TVE, T5, A3, \emptyset\}$  to watch at the beginning of day  $d$  based on the following indirect utility :

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

where  $x_{jd}^k$  represents the  $k \in \{R+, R-, L+, L-, political\}$  characteristic on channel  $j$  and day  $d$  as defined in 1.  $\sigma^k$  is a shift on characteristic  $k$  mean preferences according to unobserved preferences  $\nu_{ird}^k \sim N(0, 1)$ . 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  $\pi^k$  capture 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 seasonality.

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.

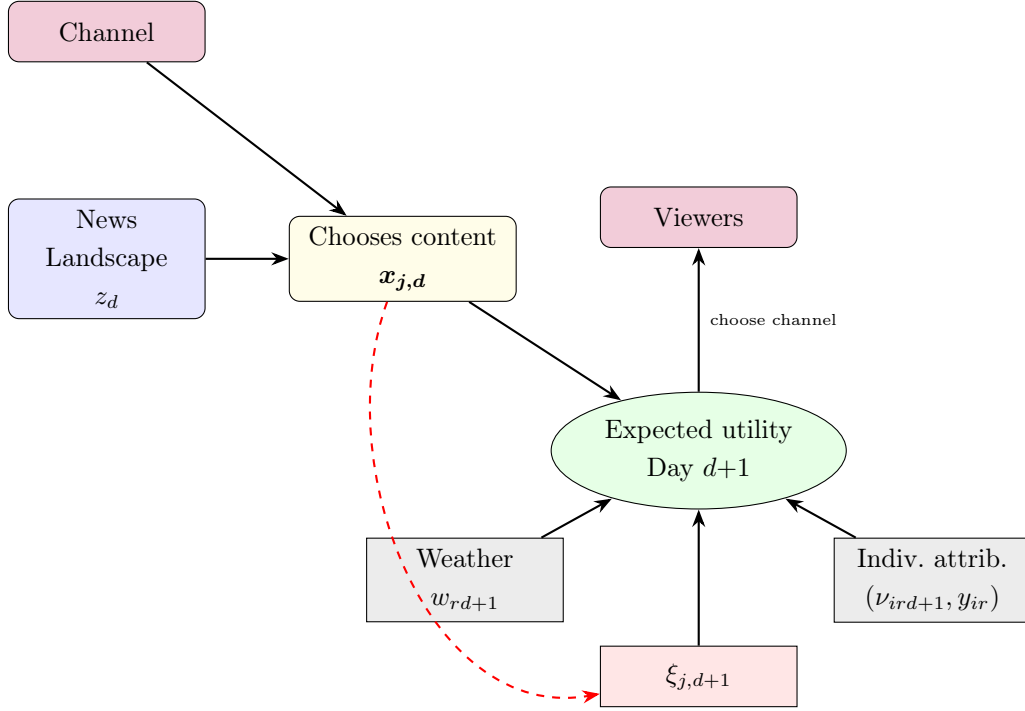
Market shares are just the integral over the individual choice indicators:

$$s_{jrd} = \int d_{ijrd}(\delta_{rd}, \mu_{ird}) d\mu_{ird} d\epsilon_{ird} \quad (3)$$

---

<sup>6</sup>The model is estimated under the *pyblp* package (Conlon and Gortmaker, 2020).

**Figure 6:** Trade-offs in the model



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. Figure 6 illustrates the trade-offs in the model. Viewers make expectations about the utility for each choice. This depends on exogenous factors such as weather, individual attributes (i.e ideology) and shocks that are unobserved to the econometrician;  $\xi$ . On the supply side, channels make their content decisions for the next day forecasting their audience tastes and the shocks that they might have the next day. Editors tweak today's rundown knowing which stories will make viewers lean in tomorrow. The red dashed arrow indicates thus the endogeneity concern  $\mathbb{E}(\mathbf{x}_{jrd-1}\xi_{jrd}) \neq \mathbf{0}$ .

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 ( $N \approx 41K$ ) 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 bias or being uninformed<sup>7</sup>.

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:

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

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 as in the TV data.

Importantly for the specification, I assume that 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. Channel would like to stick to their ideal positions (i.e those shown in 5) 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 6 and 7 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 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

---

<sup>7</sup>See collaborations for [Mediaset](#), [Atresmedia](#), and [RTVE](#).

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:

$$x_{jd}^{party+} = \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{party+}) \beta_j^{(party,+)} + \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{party-}) \beta_j^{(party,-)} + \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{political}) \delta_j + \epsilon_{jd} \quad (5)$$

$$x_{jd}^{party-} = \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{party+}) \phi_j^{(party,+)} + \sum_{party \in \{L,R\}} \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{party-}) \phi_j^{(party,-)} + \sum_{j' \in \mathcal{J}} (d(j')_j \times z_d^{political}) \rho_j + \epsilon_{jd} \quad (6)$$

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 7 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 altogether 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.

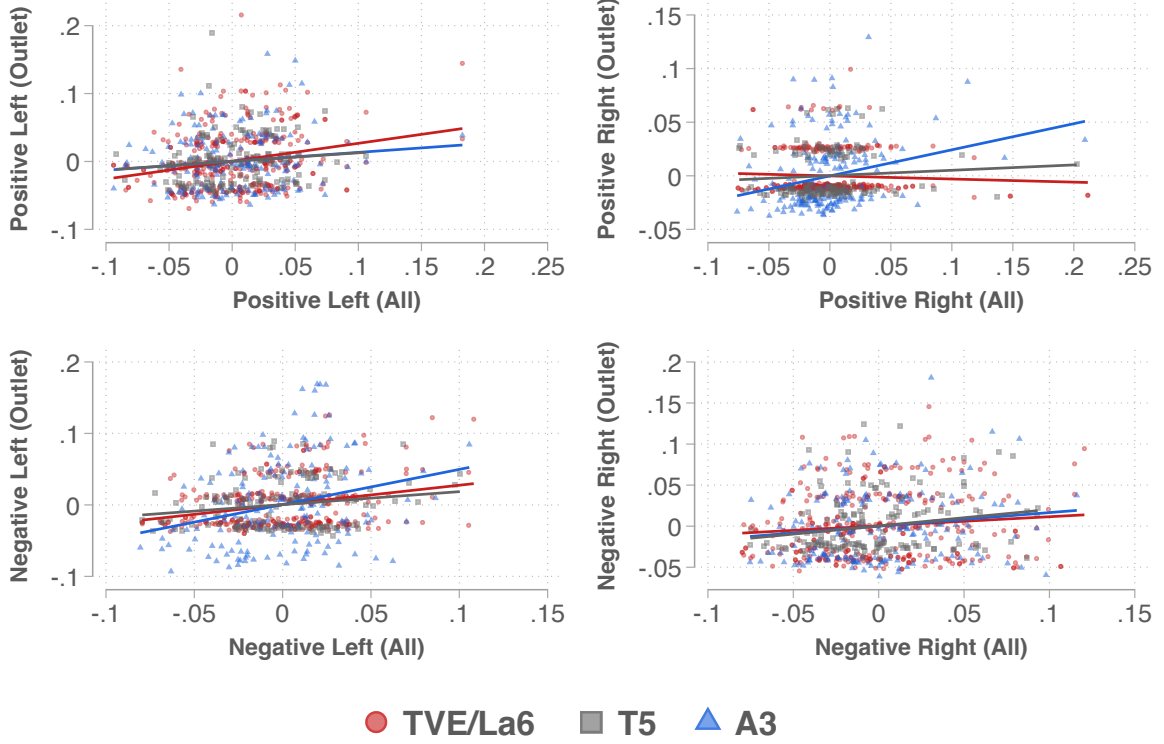
Table 1: F and t-tests for Regression Coefficients

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.



**Figure 7:** 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.

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  $k$ th 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:  $\bar{y}_r \hat{x}_{jd}^k$ .

## 6 Results

I present results of the BLP estimation for the pre-campaign and campaign periods in table 2. 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 ( $\sigma^{R-}$ ) in the population. Importantly, this dispersion is not explained by ideology heterogeneity ( $\pi$  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.

Given that increasing the positive relative minutes to a party also implies increasing the overall total minutes on politics, I compute them as:

$$\bar{\epsilon}^k = \frac{1}{J \times R \times D} \sum_j \sum_r \sum_d \left( \frac{\partial s_{jrd}}{\partial x_{jt}^k} + \frac{\partial s_{jrd}}{\partial x_{jt}^{political}} \right) \frac{x_{jd}^k}{s_{jrd}} \quad \forall k \in \{R+, R-, L+, L-\} \quad (7)$$

The estimated elasticities for right and left-markets <sup>8</sup> across tones are shown in figures 8 and 8 for the pre-campaign and campaign periods; respectively. estimate results are also shown in table ?? on the Appendix.

The polarization in preferences clearly emerges in the right pane, where it is shown how the right and left markets present opposed elasticities for the tones that opposed their ideological views.

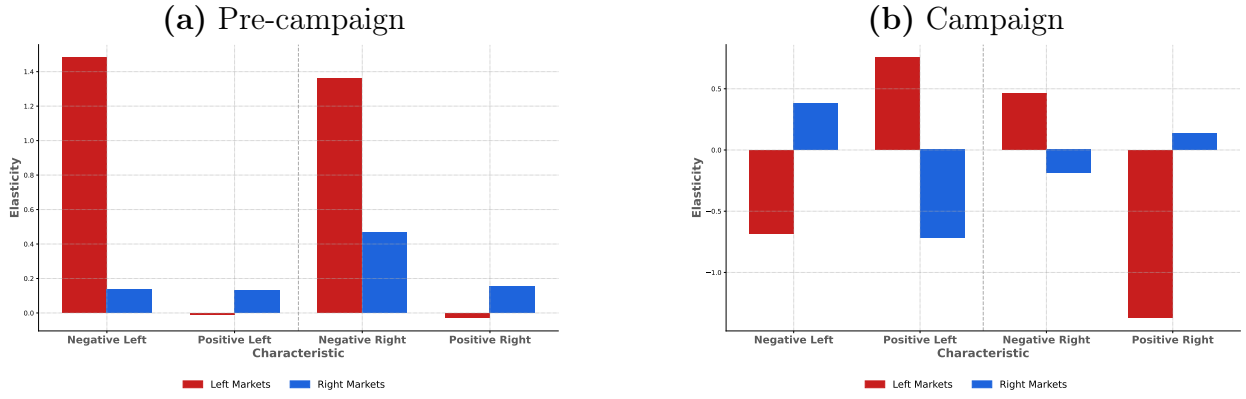
To give a size effect of this change; an increase of 1 per cent on net tone towards a party (i.e  $\bar{\epsilon}^{party+} - \bar{\epsilon}^{party-}$ ) is associated with a 0.31 per cent increase in audience share during campaign periods vs a 0.31 per cent decline in pre-campaign.

This means that an increase of 1 per cent of negative right tone during the campaign is associated with a loss of 8015 total viewers in right markets with respect to the pre-campaign period.

---

<sup>8</sup>Where a region is define as right if the proportion of right-wing voters is above the median.

**Figure 8:** Estimated Elasticities for Right and Left markets, Pre-campaign and Campaign



*Note:* Each panel shows estimated elasticities for consumer responses in right- and left-leaning markets. Panel (a) reports results for the pre-campaign period, while Panel (b) covers the campaign period.

## 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.

“Even in a news program, it’s deeply subject to the daily variation of audiences [...] 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<sup>9</sup>

Two important facts emerge from this quote. First, even though adds do not directly appear during the news slot; profitability is a key driver in their decisions. Outlets want to retain as much as audience as possible for the next slots. Second, their content choice decisions are done in a high frequency basis: Journalists meet every day and consider both the audience numbers and the events of the day in order to make their editorial lines.

Under that motivation, I model the maximization problem for a given channel  $j$  on day  $d$  is:

<sup>9</sup>Source: [https://cadenaser.com/ser/2014/12/03/television/1417630810\\_539829.html](https://cadenaser.com/ser/2014/12/03/television/1417630810_539829.html)

---


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

Channels want to retain maximum amount of audience over all the regions:  $\sum_r s_{jrd+1}(\mathbf{x}_{jd}, \mathbf{x}_{-jd}) \frac{L_r}{L}$  who has preferences over the content they offered on that day:  $\mathbf{x}_{jd}$ . I model a simple functional form  $\mathcal{C}(x_{jd}^k, z_d^k) = \frac{(x_{jd}^k)^2}{z_d^k}$  where the inverse of the available stories acts analogously to input prices. The goal is to estimate each of the parameters  $\lambda^k$  that controls for the asymmetric production costs of each type of content.

Production of content also depends on the news landscape through  $C(\cdot)$ . Under  $\lambda > 0$ , the more stories of a type there are available on a given day, i.e high  $z_d^k$ , the cheaper it is for them to produce them. These costs can have been already suggested but not estimated in previous works (Simonov and Rao, 2022). On the other hand, if channels act as news aggregators a higher volume of stories could also increase costs since they need to read and filter what they will actually show. In this case  $\lambda < 0$  would indicate an *informational filtering* type of cost with an inverse relation between the volume of stories and their production. The sign of  $\lambda$  is therefore not clear ex-ante.

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 audience reactions which precludes the need of demand shifters for the estimation.

The system of first order conditions is <sup>10</sup>:

$$\sum_r \frac{\partial s_{jrd+1}}{\partial x_{jd}^k} \frac{L_r}{L} - 2\lambda_j^k \frac{x_{jd}^k}{z_d^k} + \eta_{jd}^k = 0 \quad \forall j, k \tag{9}$$

Equation 9 can be estimated with GMM under  $\mathbb{E}(\boldsymbol{\eta}_d) = 0$ ; and it is a particular case of the *common coefficients* GMM estimation (Hayashi, 2000).

Cost parameters capture efficiency in story telling from the channels that comes from factors other than demand. Boxell and Conway (2022) show that journalist have ideology and this is an important factor of sorting across firms. If editors hire (or retain) reporters whose

---

<sup>10</sup>Notice that constrains can be ignored for estimation but should be recovered when the game is used for counterfactual analysis.

ideological leanings fit the outlet’s brand, the reporter’s ideology shows up as lower marginal effort when she covers “her” side, not as a new taste term in viewers’ utility. Outlets might also enjoy higher benefits by saving in discretion and hiring journalists that are willing to work in positions in which they can advance their careers (Baron, 2006).

The estimated parameters are shown in table 3. Content characteristics are grouped into content that favors the right and the left (top and bottom row blocks). Within each block coefficients coincide in sign for all the outlets. This might be a byproduct of the modeling assumption were every channel draws from the same wire services. Idiosyncrasies in the pool of these stories might make all channels face costs or benefits in producing them.

Table 3: Estimated Cost Parameters ( $\lambda$ ) by Channel and Content Type

	LA6	TVE	T5	A3
Characteristic				
Negative Left	-0.211	-0.345	-0.560	-0.305
Positive Right	-1.367	-1.747	-3.952	-2.380
Positive Left	0.179	0.144	3.138	0.674
Negative Right	0.076	0.144	1.601	0.299
Political	-0.005	-0.007	-0.066	-0.012

*Note:* Estimated  $\lambda$  from equation 9.

Producing both negative right and positive left content (i.e content favorable to the left) has an associated cost with positive  $\lambda$  values. The left channels La6 and TVE have lower cost of production of this type compared to the right wing channel, A3. The situation is reversed for content favorable to the right. In this case, producing this content increases the payoff of all the channels but the right wing outlet wings significantly more than the left ones. Finally, notice how the centrist Telecinco (T5) dominates the absolute values in every row.

## 8 Counterfactuals

Across Europe, broadcast regulators increasingly impose proportional-air-time rules to curb partisan imbalances in election coverage: France’s ARCOM stopwatches every political segment to enforce strict minute-for-minute equality during the “période officielle”; Italy’s Par Condicio law compels both public and private stations to “balance and compare” exposure for all lists; and Germany grants each registered party free television slots whose length scales with its prior vote share. Spain, by contrast, still relies on a voluntary pluralism. The public

---

network TVE issues a proportional airtime manifesto while commercial groups—Atresmedia and Mediaset—entirely unconstrained.

My model estimates allow to test what would be the implications of enforcing such policy to all the broadcasters. I take the proportion of votes in the past general elections for each party and estimate the new equilibrium under the following problem:

$$\begin{aligned}
& \max_{\{\mathbf{x}_{jd}\}} \left\{ \sum_r s_{jrd+1}(\mathbf{x}_{jd}, \mathbf{x}_{-jd}) \frac{L_r}{L} - \sum_k \lambda^k \mathcal{C}(x_{jd}^k, z_d^k) + \boldsymbol{\eta}_{jd} \mathbf{x}_{jd} \right\} \\
& s.t. \quad \frac{x_{jd}^{R+} + x_{jd}^{R-}}{x_{jd}^{political}} = vote^R \\
& \quad \frac{x_{jd}^{L+} + x_{jd}^{L-}}{x_{jd}^{political}} = vote^L \\
& \quad x_{jd}^k \geq 0 \quad \forall k
\end{aligned} \tag{10}$$

where I assume that time to a political party can only be discerned if that story has some specific slant and that all outlets must devote the proportion of political time given by previous proportions of votes ( $vote^R, vote^L$ ) (i.e constraints bind).

Table 2: BLP Estimation Results with Standard Errors

Coefficient	Parameter	Estimate	Std. Error
Pre-campaign			
Positive Left	$\beta^{L+}$	-16.90	(11.73)
Positive Right	$\beta^{R+}$	-25.53	(31.03)
Negative Left	$\beta^{L-}$	28.29	(25.89)
Negative Right	$\beta^{R-}$	56.44**	(28.45)
Political	$\beta^{political}$	11.49***	(4.28)
Weather	$\gamma$	0.00	(0.03)
Positive Left	$\sigma^{L+}$	0.44	(296.90)
Positive Right	$\sigma^{R+}$	0.68	(715.26)
Negative Left	$\sigma^{L-}$	12.64	(8.68)
Negative Right	$\sigma^{R-}$	21.99	(13.83)
Political	$\sigma^{political}$	0.00	(46.79)
Right-Wing $\times$ Positive Left	$\pi^{L+}$	46.26	(48.70)
Right-Wing $\times$ Positive Right	$\pi^{R+}$	77.81	(95.92)
Right-Wing $\times$ Negative Left	$\pi^{L-}$	-70.09	(57.97)
Right-Wing $\times$ Negative Right	$\pi^{R-}$	-103.71	(97.76)
Right-Wing $\times$ Political	$\pi^{political}$	-25.17**	(12.59)
Campaign			
Positive Left	$\beta^{L+}$	134.16**	(65.57)
Positive Right	$\beta^{R+}$	-129.79***	(48.43)
Negative Left	$\beta^{L-}$	-104.84**	(41.32)
Negative Right	$\beta^{R-}$	92.62**	(41.63)
Political	$\beta^{political}$	-6.43**	(3.26)
Weather	$\gamma$	0.00	(0.01)
Positive Left	$\sigma^{L+}$	0.00	(320.36)
Positive Right	$\sigma^{R+}$	19.25*	(10.89)
Negative Left	$\sigma^{L-}$	0.00	(134.33)
Negative Right	$\sigma^{R-}$	0.02	(73.18)
Political	$\sigma^{political}$	0.00	(19.55)
Right-Wing $\times$ Positive Left	$\pi^{L+}$	-421.99**	(176.15)
Right-Wing $\times$ Positive Right	$\pi^{R+}$	334.94***	(127.41)
Right-Wing $\times$ Negative Left	$\pi^{L-}$	288.65**	(116.22)
Right-Wing $\times$ Negative Right	$\pi^{R-}$	-280.33***	(104.44)
Right-Wing $\times$ Political	$\pi^{political}$	19.34**	(8.93)

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 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. Finally, I model the competition game into channels' daily content production and estimate their cost structures. Right wing media enjoy higher benefits/costs on producing content that goes against their editorial line.



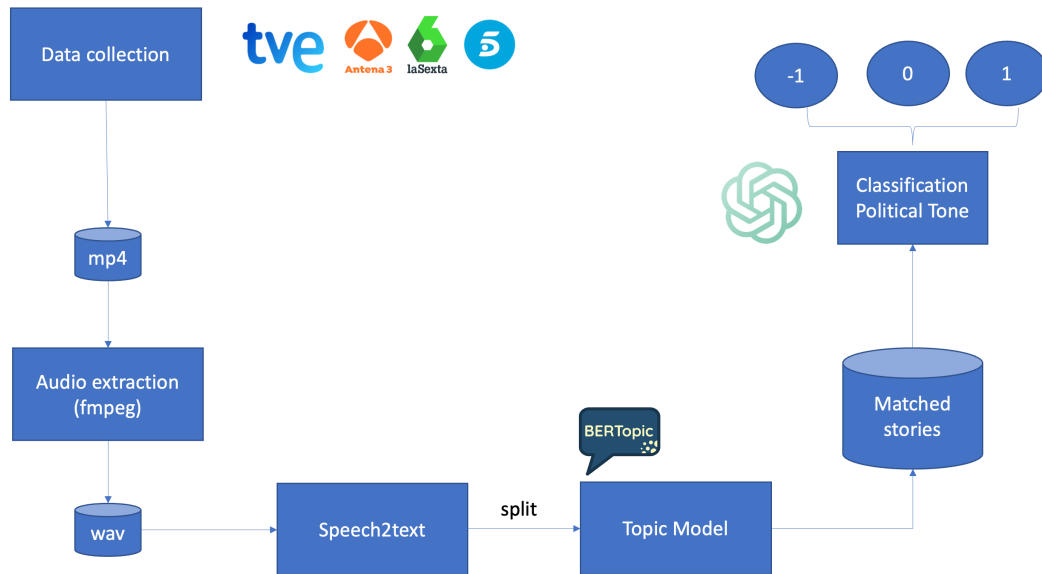
---

## 10 Appendix

### 10.1 Figures

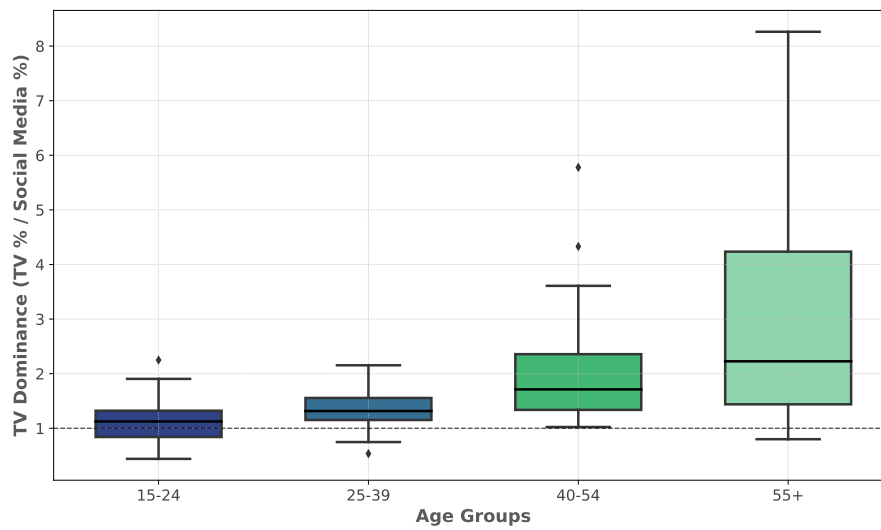
Pipeline:

**Figure 9:** Pipeline for content downloading and classification



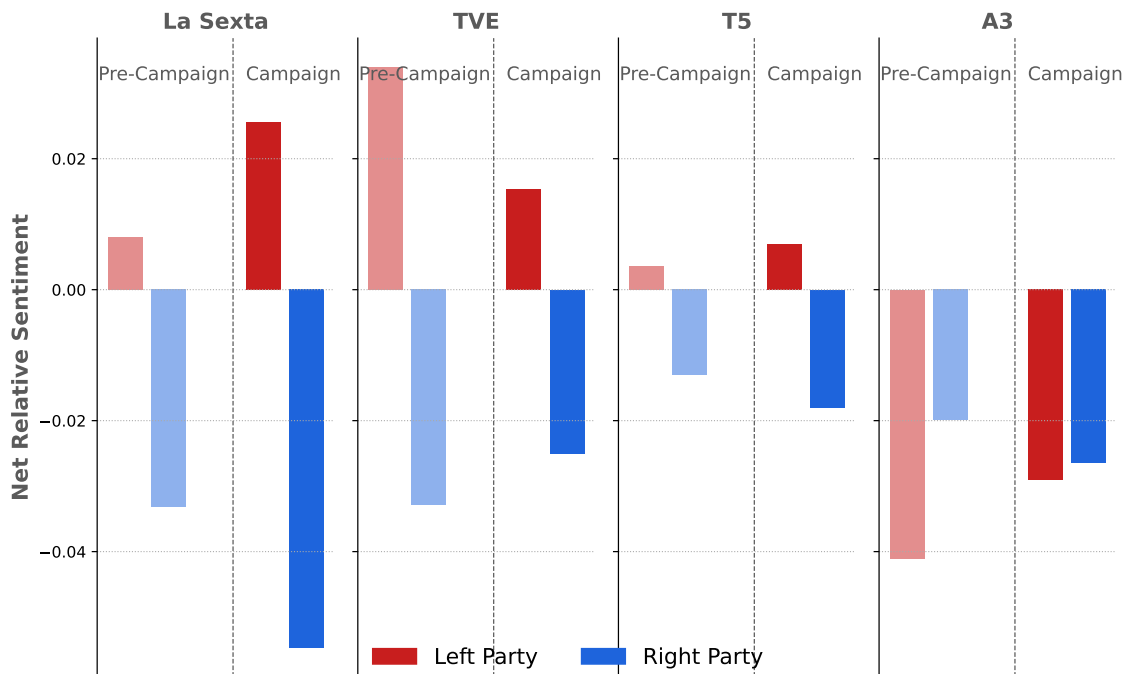
Notes: Pipeline for the text downloading. First videos are downloaded daily from the main TV channels. Google engine is used to convert the audio to text. I then split the stories by minute and user BERTopic to classify and match them. Finally, ChatGPT4 is used to classify political tone.

**Figure 10:** 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.

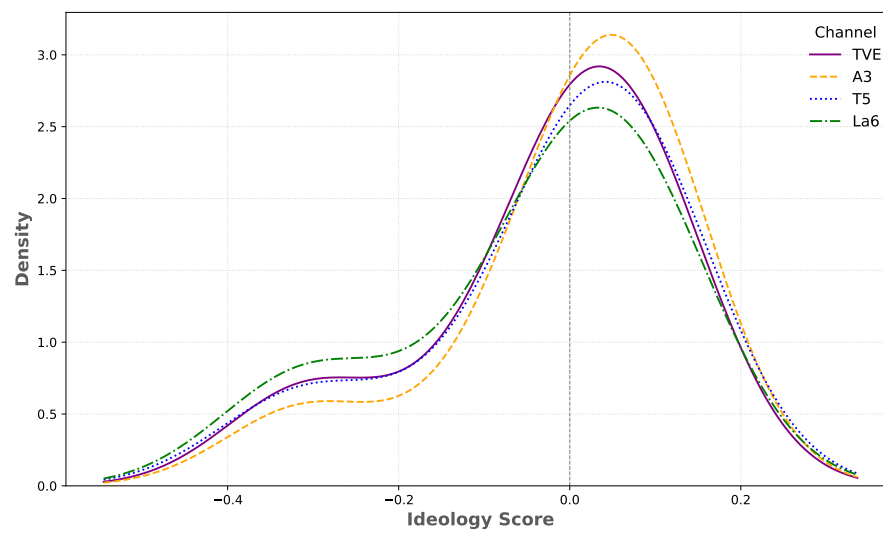
**Figure 11:** Net sentiment across channels and parties pre and during campaign



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 12:** 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 data and weighted by channels' share of audience for each autonomous region.

## 10.2 Tables

Table 4: Top 5 Topics by party-tone category on Agencia EFE after BERTopic.

Topic Words	Count
<b>Positive Right</b>	
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
<b>Negative Right</b>	
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
<b>Positive Left</b>	
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
<b>Negative Left</b>	
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 5: 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.

---

### 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	0.037
Antena 3	0.184
Telecinco	0.037
La Sexta	0.01

---

Table 6: 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.

---

### 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	0.067
La Sexta	0.148

---

Table 7: 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.

Positive Right	Negative Right
secretario general vox	gürtel servicio nacional
vox ignacio garriga	juicio gürtel servicio
general vox ignacio	valenciana francisco camps
núñez feijóo llamado	gürtel juicio gürtel
candidato pp elecciones	expresidente generalitat valenciana
partidos vox madrid	generalitat valenciana francisco
josé sáenz buruaga	vicesecretaria general psoe
núñez feijóo pedido	juicio gürtel madrid
maría josé sáenz	reforma ley aborto
may presidente pp	expresidente valenciano francisco

Table 8: Top trigrams for Positive Right and Negative Right on the Agencia EFE dataset using the ChatGPT categories.

Positive Left	Negative Left
agricultura pesca alimentación	tribunales ere sevilla
ministro agricultura pesca	derechos sociales ione
dic presidente gobierno	sociales ione belarra
pesca alimentación luis	junta andalucía josé
alimentación luis planas	andalucía josé antonio
ene presidente gobierno	aplicación ley solo
vicesecretaria general psoe	expresidente junta andalucía
consejo ministros aprobado	caso mediador palmas
ministra asuntos económicos	mediador palmas gran
vicepresidenta primera ministra	núñez feijóo acusado

Table 9: Top trigrams for Positive Left and Negative Left on the Agencia EFE dataset using the ChatGPT categories.

Table 10: Estimated Elasticities by Left and Right markets

Characteristic	Pre-campaign		Campaign	
	Left Market	Right Market	Left Market	Right Market
Negative Left	1.481	0.136	-0.681	0.380
Positive Left	-0.011	0.129	0.754	-0.716
Negative Right	1.358	0.469	0.462	-0.182
Positive Right	-0.029	0.152	-1.371	0.133

*Note:* The table shows the average elasticities for each characteristic by right and left markets as defined in equation 7. Right markets are defined as regions with proportion of right wing voters above the median.

---

### 10.3 Examples stories and classification

Story	Channel	Date	Sentiment
A survey of 6,000 teenagers in Madrid schools reveals that 26% do not identify as heterosexual, though fear of expressing it publicly is growing. Trans students represent 3%. The Minister of Equality has called for commitment against harassment. She emphasized the importance of addressing not only positive progress but also areas where work remains, condemning normalization of hate speech...	TVE	2023-04-14	positive UP
In Mallorca, Unidas Podemos claims the elections are a referendum on the right to housing. They take credit for the new housing law and accuse their government partners of delaying its approval over the legislative term...	A3	2023-05-15	positive UP
Pedro Sánchez will pause his electoral campaign to attend the NATO Summit in Vilnius, following Spain's assumption of the EU Council presidency. U.S. President Joe Biden will also attend after a stop in London to meet King Charles III.	A3	2023-07-09	positive PSOE
Prime Minister Pedro Sánchez visited the command center for those affected by a wildfire, linking it to climate change and urging serious action. He announced a cabinet reshuffle, naming new ministers for Health, and Industry, Trade, and Tourism.	TVE	2023-03-27	positive PSOE
Regarding the Penal Code reform, the Popular Party leader called for reinstating the sedition charge, criticizing Pedro Sánchez's decisions as ridiculous and opportunistic.	TVE	2023-02-14	positive PP

---

PP leader Feijóo criticized the PSOE for aligning with Podemos, suggesting a streamlined government structure with 13-14 ministries compared to the current 22. He equated PSOE's approach to that of Podemos while Pedro Sánchez likened PP to Vox.	A3	2023-06-01	positive PP
The new PP government in the Balearic Islands abolished inheritance and donation taxes for direct relatives, reduced taxes for other relatives, and lowered property transfer taxes for young first-time buyers.	A3	2023-07-18	positive PP
Vox is key in provincial capitals like Valladolid, where their alliance could secure an absolute majority. Talks are ongoing in regional governments, with Azcón meeting Vox representatives to negotiate a coalition for Aragón.	LA6	2023-06-13	positive Vox
Vox filed its second motion of no-confidence against the government, this time with a candidate who is not the party leader.	A3	2023-02-27	positive Vox
In Extremadura, the Socialist Party risks losing its majority, while the PP and Vox could enter the regional parliament for the first time, potentially tipping the balance after the elections.	T5	2023-05-23	positive Vox
On Thursday, the commission for the reform of the law begins. A reform that creates significant division between the Socialist Party and Podemos. Podemos is also facing another challenge: negotiations with Yolanda Díaz to run together in the elections. - They are still searching for a formula to run together in the general elections. And since no agreement has been reached yet, Belarra is asking Yolanda Díaz today to campaign for Podemos in the May 28 elections.	A3	2023-04-09	negative UP



---

Here in Spain, a day after the launch of her platform, Yolanda Díaz continues exchanging criticisms with Podemos. The Podemos party did not attend, highlighting the division to the left of the PSOE. Díaz urges them to explain their absence while asserting that a "Sumar" without Podemos would not be a failure. This diagnosis contrasts sharply with that of Podemos. It remains to be seen whether, in the end, Sumar decides to proceed without Podemos.	T5	2023-04-03	negative UP
The alleged extortion scheme involving former Socialist deputy Juan Bernardo Fuentes, who is under investigation. The PSOE highlights the swift action taken by the party to suspend his membership and force him to relinquish his seat. We need explanations, and we need them now because, with each passing day, what we learn only adds to the shame, and it keeps getting worse.	TVE	2023-02-27	negative PSOE
The weekend arrives after a few intense days for the governing coalition. The insults from Podemos towards the PSOE continue over their reliance on the PP to reform the "Only Yes is Yes" law. However, nothing changes, and now everyone is focusing on the campaign for the May 28 elections.	A3	2023-04-21	negative PSOE
"We're going in, and we're stopping the machines." Residents of San Fernando de Henares, in Madrid, temporarily halted Metro construction works that are damaging their homes. They accuse Ayuso's government of not listening to them. Workers are injecting cement to stabilize the ground, which has shifted due to the subway extension—a measure that is cracking their walls.	LA6	2023-01-05	negative PP

---

The Supreme Court upholds the sentences handed down by the National Court to the ringleaders of the Gürtel Scheme for the embezzlement of public funds during the Pope's visit to Valencia in 2006. Francisco Correa has been sentenced to 13 years in prison for rigging contracts, Pablo Crespo to 15 years, and "El Bigotes" to 6 years.	A3	2023-04-10	negative PP
But this is the strategy devised by the Madrid PP, led by Ayuso, and endorsed by the national PP under Feijóo: to raise the level of tension in the discourse to take votes away from Vox in the Community of Madrid. This is part of Ayuso's attempt to secure an absolute majority, which, according to polls, is within reach but not guaranteed.	A3	2023-05-18	negative VOX
The results could be excellent. That said, like my colleagues, I must tell you that nothing is certain yet. Sumar is calling for votes to avoid a return to black-and-white. Yolanda Díaz is rallying women against the prospect of Abascal becoming vice president under Feijóo and predicts a return to an austerity-driven Spain with a right-wing government.	LA6	2023-07-15	negative VOX

Table 11: The table shows examples of fragments of stories associated with different tone for each party according to the Chat GPT classification. Stories has been translated to English with the help of ChatGPT.

## 10.4 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 12 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 moneros
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 12: 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.

I 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.

---

## 10.5 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 average 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 13 shows the mean and standard deviation of the classification output for 100 rounds of classification.

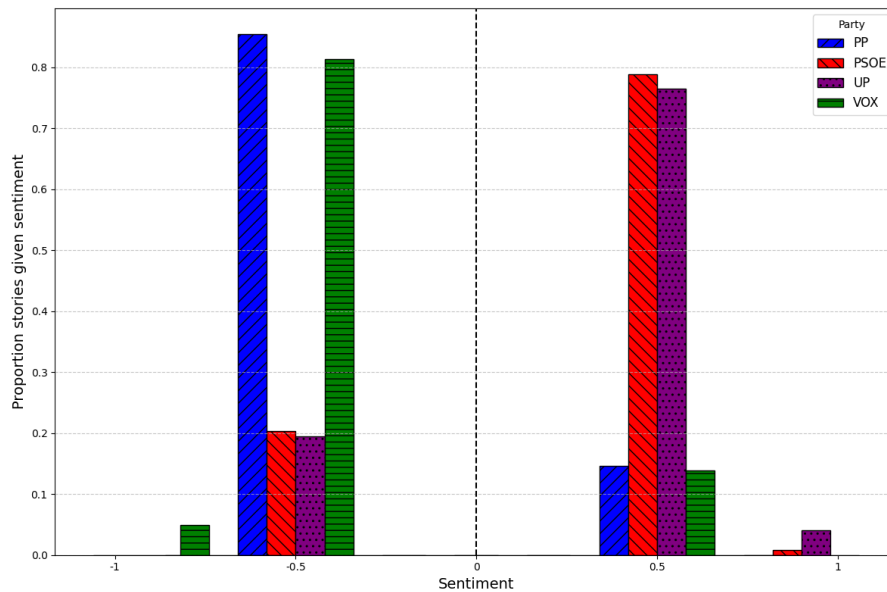
Results of the final classification for the non neutral stories are shown in figure 13. Each bar represents the percentage of stories of that given sentiment associated to each political party. We can see that the classification reserved the extreme values 1 and -1 for few stories and focused on the -0.5 and 0.5 values.

Table 13: 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.

Figure 13: 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.

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.

$$\text{Score}(w) = \ln \left( \frac{\text{freq}(w, \text{Left}) + \alpha}{\text{Total}_{\text{Left}} + \alpha} \right) - \ln \left( \frac{\text{freq}(w, \text{Right}) + \alpha}{\text{Total}_{\text{Right}} + \alpha} \right), \quad (11)$$

- $\text{freq}(w, \text{Left})$  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
- $\alpha$  is a small smoothing parameter.

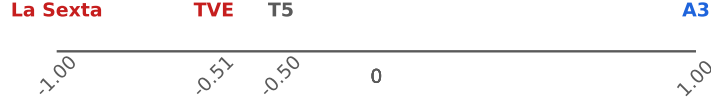
I select the value of alpha that maximizes accuracy of label prediction in the congress dataset;  $\alpha = 0.9$ . Words with high positive scores are used disproportionately in left-leaning speeches, while those with high negative scores are more characteristic of right-leaning speeches. I rank all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

jornaleros orihuela telemático billy  
renovar ciudadanas republic  
**astur**ies neoliberales  
paisaje vecinas moitas emitidos  
privatizaciones  
**telemáticos**  
Comparecerá minoritarias  
artista audios elimino concertación macrogranjias  
oreg efectuada asturiana cloacas  
mecenazgo asturianas aprem mais sara  
**neoliberal** chillas sintientes  
interiores juvenit interino  
ludopatía oligopolio cedaw  
apuestas avanzados guarín versos superpagadores contrarios

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 16:** Normalized Ideology Scores



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

## 10.6 Instrument validity

The instrumental variable approach outlined in section 5.1 relies on the assumption that the news landscape, proxied by the pool of stories, is a pure supply shifter. This assumption would be violated if viewers exhibit some anticipation behavior. Viewers naturally know that, for instance, a scandal occurred; the assumption is that they cannot see is the volume and polarity of the raw wire material that the channels count on. As a diagnostic, I regress minute-0 audience shares on the available pool of stories on the same day. In this sense, these results serve as a falsification test where, if there was anticipation behavior, initial audience would tune into different channels according to the news landscape of that given day.

$$\ln \left( \frac{q_{jt}}{L_r - Q_{jt}} \right) = \beta_0 + \sum_{p \in \{L, R\}} (\rho^{p+} z_d^{p+} + \rho^{p-} z_d^{p-}) + \phi z_d^{political} + \gamma_{dow} + \gamma_r + \epsilon_{jt} \quad (12)$$

Table 14 show the estimated coefficients from equation 12 where each column corresponds to a separate outlet regression and I condition on the minute zero audience. Equation 12 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.

## 10.7 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

---

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

---

Robust standard errors in parentheses

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

Table 14: The table shows the results from the estimation of [12](#) 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.

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 17](#)), which is the area where the output is expected to appear.

---

**Figure 17:** 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.

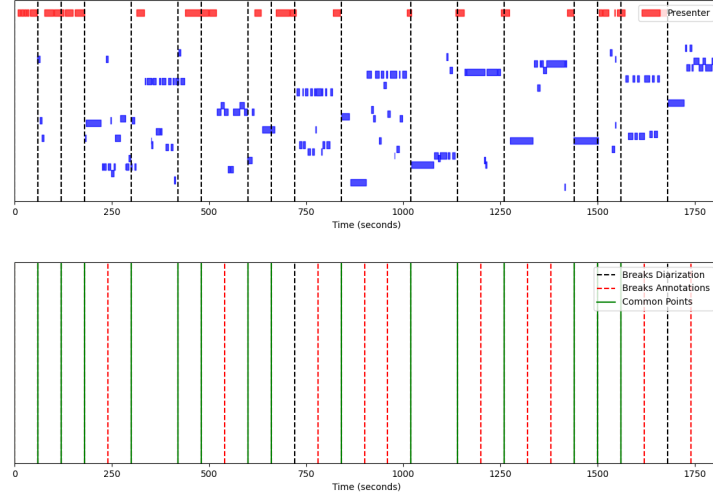
### Speaker diarization

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 18 illustrates this procedure and the comparison with the manually annotated labels for an example day-channel.



---

**Figure 18:** 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

- Atil, B., Chittams, A., Fu, L., Ture, F., Xu, L., and Baldwin, B. (2024). Llm stability: A detailed analysis with some surprises. *CoRR*, abs/2408.04667.
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Fallin Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.
- Baron, D. P. (2006). Persistent media bias. *Journal of Public Economics*, 90(1–2):1–36.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.
- Boxell, L. and Conway, J. (2022). Journalist ideology and the production of news: Evidence from movers. SSRN Working Paper No. 4295587.
- 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.
- Edelman (2023). Edelman trust barometer 2023: Global report.
- 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 Stone, D. F. (2014). Media bias in the marketplace: Theory. NBER Working Paper 19880, National Bureau of Economic Research.
- 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.
- Longuet-Marx, N. (2025). Party lines or voter preferences? explaining political realignment. Presented at the Spring 2025 IOG-BFI Conference, March 28–30, 2025.
- Newman, N., Fletcher, R., Schulz, A., Andl, S., Robertson, C. T., and Nielsen, R. K. (2023). Digital news report 2023. Accessed: 2025-05-13.
- Parliament, E. (2024). Audiovisual and media policy.
- Peterson, E. (2018). Echo chambers and partisan polarization: Evidence from the 2016 presidential campaign. Working paper.
- 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.
- Schneider-Strawczynski, S. and Valette, J. (2025). Media coverage of immigration and the polarization of attitudes. *American Economic Journal: Applied Economics*, 17(1):337–368.
- 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*.
- Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020). Political effects of the internet and social media. *Annual Review of Economics*, 12(1):415–438.