

Media-Polarisation in Denmark

Beyond the Headlines: Textual and Visual Content
from a Decade of News

Knud Anton Højmark Bremholm
Christian Andersen Mølby

ABSTRACT

We study media polarisation in the two Danish online news outlets DR and TV2 based on article text and lead images. By combining computer vision with natural language processing, we extract features from the article images. In combination with machine learning techniques and statistical extraction methods this allows us to distil a wide range of tokens from 168.197 articles published between 2015 and 2024. From a total of 11.954.367 tokens, we obtain unbiased estimates of polarisation using a leave-out estimator and compare model performance to maximum likelihood estimates. We document significant estimates of polarisation in all but 8 out of 260 two-week periods. The average level of polarisation is 0,534, indicating a 53 pct. probability of a neutral observer with full information correctly identifying the source of an article, given a single, randomly drawn token. This level is comparable to previous findings from American online news. We reject that polarisation increases during the studied period and find a regime of persistent moderation in polarisation over the last three years. Having established the existence of polarisation in Danish media, we test whether it follows an electoral cycle. Our findings show that polarisation decreases in the lead-up to general elections.

1 INTRODUCTION

“The political media is biased, but not toward the left or right so much as toward loud, outrageous, colorful, inspirational, confrontational.”

— Ezra Klein, *Why We’re Polarized* (2020)

High polarisation in a society can have adverse effects, for instance, it poses a potential threat to the effective functioning of democratic institutions ([VIVE, 2022](#)). In extension, increases in polarisation "... may have important consequences, including reducing the efficacy of government, increasing the homophily of social groups, and altering economic decisions" ([Boxell et al., 2020](#))¹. These aspects emphasise the importance of assessing polarisation and how it develops over time.

Acknowledgements: We thank Københavns Universitet and KUB-datalab for providing access to the extensive computational resources employed in the feature extraction and analysis. Further, extend our gratitude to Anne Sofie Beck Knudsen for valuable feedback and guidance.

¹Boxell et al. ground these consequences in findings from [Hetherington and Rudolph \(2015\)](#), [Iyengar et al. \(2012\)](#), [Iyengar et al. \(2019\)](#), and [Gift and Gift \(2015\)](#)

In the United States, a substantial body of research has documented how polarisation has increased over recent decades ([Boxell et al., 2020](#); [Gentzkow et al., 2019](#)). As an example of this, [Andris et al. \(2015\)](#) demonstrate that bipartisan cooperation has steadily declined in the US Congress over the past 60 years. Despite the historically consensus-oriented political culture in Denmark, this tendency may translate to a Danish context ([Hjorth et al., 2019](#)). Therefore, it raises the central question that motivates our study: Is polarisation increasing in Denmark?

This paper concerns polarisation in media content, and we define polarisation as a measure of divergence of how media cover similar material. Our approach estimates polarisation in both text and images, based on 168.197 article-image pairs from DR.dk and TV2.dk, during 2015–2024. Drawing on recent advances in computer vision, natural language processing, and neural networks we develop state-of-the-art scene tagging, complemented by classical machine learning for facial attributes and recognition. Combined with keyword extraction and sentiment scores, this yields a data foundation unprecedentedly rich in Danish media studies. For estimation, we theoretically evaluate three candidate estimators and identify the leave-out estimator as the most suitable. We apply it estimate actual polarisation and a control polarisation based on random assignment of media to the articles. We estimate significant polarisation in 252 out of 260 two-week periods and we go on to show econometrically how polarisation fluctuate with elections.

2 LITERARY CONTEXT

The advances in artificial intelligence and machine learning enables the analysis of unconventional data sources, extending the range of empirical contexts for econometric analysis. This paper contributes to the emerging field of empirical polarisation studies, situated at the intersection of econometrics, social data science, and media studies.

Research on polarisation often centres on affective polarisation. This includes polarisation in sympathies toward political parties, stereotypes of opponents, or social distance to groups with different opinions. It is often quantified through surveys, such as whether respondents accept a neighbour with a different political affiliation ([Hjorth et al., 2019](#)). The study shows signs of social distance to political opponents in Denmark, where the phenomenon previously has been documented mainly in the U.S. [Boxell et al. \(2020\)](#) analyse affective polarisation across 12 OECD countries, documenting increases in five (including Denmark) over four decades, driven partly by ideological distance between parties and shifts in news consumption, while [Kingzette et al. \(2021\)](#) argue that high affective polarisation risks undermining democratic norms. [VIVE \(2022\)](#) summarises existing Danish studies, predominantly of affective polarisation, and conclude that while

some polarisation exists it appears to be a fundamental democratic condition.

Our paper examines polarisation in news rather than affective polarisation. The two are related, since affective polarisation fosters echo-chambers that news media may reinforce. In the U.S., [Kim et al. \(2022\)](#) and [Ash et al. \(2023\)](#) show that cable news outlets like Fox News have contributed to increased polarisation. This connects to the concept of the Overton window, describing the boundaries of acceptable speech ([Russell, 2006](#); [Mackinac Center for Public Policy, 2023](#)). This framework is useful for understanding polarisation in media coverage ([The New York Times, 2019](#); [Bursztyn et al., 2020](#)).

In a Danish context [Enevoldsen and Hansen \(2017\)](#), use sentiment analysis of 360 articles from Berlingske and Information, find significant interactions between newspapers and parties. They highlight the potential of web-mined datasets. Our paper responds to this call, combining large-scale data and econometric methods to advance the study of media polarisation. We draw on two influential methodological contributions outlined below.

The first is [Caprini \(2024\)](#), who study visual bias in U.S. news media and document higher polarisation in images than in text. Using news shared on Twitter (2019–2020) Caprini constructs a “visual vocabulary” and estimates polarisation via the leave-out estimator ([Gentzkow et al., 2019](#)). The conclusions of the paper highlight the large degree of partisanship in the American news sphere, with a polarisation in the range between 0,516 and 0,534, with a mean of 0,525. We provide an intuitive interpretation of these measures and relate them to our findings in the subsequent analysis. The paper finds strong visual partisanship motivating our focus on images².

The second is [Gentzkow et al. \(2019\)](#), who originally developed methods to measure group differences when the choice set is high-dimensional and applied to U.S. congressional speech. They propose the leave-out and penalised estimators to correct finite-sample bias, finding stable polarisation historically but sharply rising since the 1990s.

Together, these studies provide a framework for quantifying media content and estimating polarisation. Our contribution differs from the papers above in using multi-dimensional data, drastically increasing the time horizon, and following up with time-series econometric analysis of estimated polarisation. Finally, the implementation of polarisation estimators on multi-dimensional data and the extensive feature extraction it requires constitute in itself a methodological advancement in the field. We have thus

²In [Mølby and Bremholm \(2025b\)](#) we rely on [Ash et al. \(2022\)](#), who analyse New York Times and Fox News over 20 years using computer vision and NLP to document differences in representation of gender and ethnicities. Such systematic differences highlight dimensions along which Danish media also diverge, but we reserve this investigation to [Mølby and Bremholm \(2025b\)](#).

made our code publicly available ([Mølby and Bremholm, 2025a](#)). The code is written using [Python Software Foundation \(2025\)](#) and R ([R Core Team, 2025](#))³.

Theoretically, we expect some polarisation in the news coverage due to market powers, see appendix 7.2. and [Thomas J. Nechyba \(2018\)](#). By differentiating the news coverage, the outlets create distinct brands, potentially creating brand-loyalty, increasing the medias' incentives to reinforce deviating editorial lines. We theorise that these micro-economic drivers of polarisation result in a natural level of polarisation in Danish media. This leads us to formulate the overarching working hypothesis that there is polarisation in Danish media. This is supported by the findings in [VIVE \(2022\)](#) and [Boxell et al. \(2020\)](#), as outlined above. Further, we theorise that polarisation, through expansions in the Overton window, might be affected by specific events, as previously suggested by [Caprini \(2023\)](#). In this paper we specifically investigate the impact of elections, relating media studies to political studies. We refer to [Mølby and Bremholm \(2025b\)](#) for a complete treatment of the impact of non-recurring contemporaneous events of polarisation.

3 DATA FOUNDATION

The following section contains a detailed description of the extensive methods applied in collecting the dataset, and the advanced computational resources the process utilises.

Data sources: The analysis focuses on two major Danish online news outlets: DR.dk and TV2.dk ([DR, 2025](#); [TV2, 2025](#)). We collect all news articles and leading images from 01/01/2015 to 31/12/2024. Online news is chosen for its multi-dimensionality, allowing extraction of characteristics from both text and images. We exclude news types lacking a traditional composition of article text with an accompanying lead image and note that online news may differ from print in content and form, with a bias toward breaking news and live coverage. Online publication is necessary, as extraction from print archives is infeasible⁴. For both DR and TV2, coverage before 2015 is sparse and inconsistent in format, motivating our temporal cut-off. Methodological constraints, including the dimensionality of our polarisation estimator, further restrict the analysis to two outlets⁵. We therefore select DR and TV2, which share public service obligations and adhere to traditional article templates. DR and TV2 are central to daily news in Denmark, each with a large readership ([Slots- og Kulturstyrelsen, 2017](#); [Journalisten, 2021](#); [Danmarks Radio, 2025](#)). Popular perceptions hold DR as more left-leaning than TV2 ([Journalisten, 2010](#); [Berlingske, 2010](#)). This combination hints that, while the outlets appear somewhat similar, each media has a distinct editorial line, which they follow.

³See appendix 7.1 for full list of R packages.

⁴e.g. [www.infomedia.dk](#) where images are often unavailable.

⁵A polarisation measure with a higher dimensionality than 2 bears little intuitive meaning.

Data collection: Utilising Internet Archive’s Wayback Machine, we construct a web scraper that parses historical front pages for both news outlets ([Internet Archive, 2025](#)). It outputs links to all relevant articles for each day in the aforementioned time-horizon. We exclude articles from, e.g., live-blogs, articles without a leading main image, and all articles not categorised as news, such as sports. Appendix 7.3. outlines the full set of criteria determining relevancy⁶. Next, we build a scraper targeting DR.dk and TV2.dk. For each historical article link, we pull the corresponding web page, compiling a dataset with metadata on outlet, section, headline, date, full text, and a unique leading image. Outlet sections vary over time and overlap only partly, so we align and categorise them as described in appendix 7.4. In total, this yields 168.197 unique articles with images.

3.1 FEATURE EXTRACTION

We extract multiple features from the main image and the article text with the aim of creating a set of tokens that represent the text and image content for each article. Here, we distinguish between the list of extracted features and the resulting set of tokens. We employ three specific text- and image- analysis tools to extract the features: Scene tagging, Face recognition, and Keyword extraction. All features are compiled and formatted to construct a dataset comprising the full set of tokens from all articles.



Scene tags:

Police	Protest	Riot	Emergency	Law enforcement
Crowd control	Public safety	Nighttime	Urban Environment	Civil unrest

Figure 1: Example of object detection and scene tagging

Scene tagging: To analyse how images contribute to polarisation, we rely on recent advances in computer vision that combine convolutional neural networks (CNNs) with large language models (LLMs). At the frontier lies unsupervised scene tagging, which enables zero-shot object and scene recognition without requiring labelled training data or predefined categories ([Radford et al., 2021](#); [OpenAI, 2021](#)). We build a programme that implements one of the most advanced open-source models, LLaVa:7B, accessed via the

⁶See [Mølby and Bremholm \(2025b\)](#) for a detailed description of the data collection and its limits.

Ollama module (Liu et al., 2024; Ollama, 2024; Liu et al., 2023). This method simultaneously detects object placement, recognises objects and scenes, and uses a generative component to assign and prioritise labels (Zheng et al., 2023). This constitutes a considerable technological leap, but it comes with high computational costs: processing a single year of images takes up to three days, and the entire corpus approaches 700 hours of runtime. Despite this, the scalability across image types and the precision of the resulting tags make it uniquely suitable for our purposes. The programme ultimately returns a structured set of features per image, illustrated in figure 1. We align, format, and translate the features to allow for possible overlaps between tokens from images and text (Finlay and Argos Translate, 2025). Finally, we apply a lemmatizer to strip the tokens to their roots, ensuring that we can map tokens when grammar might differ (Lind, 2021).

Face recognition and facial attributes: While our unsupervised scene tagging approach represents the frontier of computer vision, we complement it with classical machine learning methods combined with CNNs to quantify facial characteristics in images.



Facial analysis:

Name:	(1) Marie Bjerre (V)	(2) Jakob Engel-Schmidt (M)	(3) Thomas Danielsen (V)
Distance:	0,38	0,13	0,17
Dominant emotion:	Happy	Happy	Neutral
Gender:	Woman	Man	Man
Age:	27	29	38
Race:	White	White	White
Name:	(4) Lars Løkke Rasmussen (M)	(5) Jakob Ellemann-Jensen (V)	(6) Sophie Løhde (V)
Distance:	0,16	0,08	0,33
Dominant emotion:	Sad	Neutral	Happy
Gender:	Man	Man	Woman
Age:	48	41	33
Race:	White	White	White
Name:	(7) Mette Frederiksen (S)	(8) Morten Bødskov (S)	(9) Mathias Tesfaye (S)
Distance:	0,19	0,24	0,23
Dominant emotion:	Happy	Neutral	Neutral
Gender:	Woman	Man	Man
Age:	34	41	43
Race:	White	White	White

Figure 2: Example of face detection, alignment, and facial analysis

Using advanced open-source components, we build a programme for facial analysis. First, we apply the single-shot multilevel detection model RetinaFace via the DeepFace module which with high precision locates faces (Deng et al., 2019; Serengil, 2024). This alignment directs the face recognition model Facenet512, pre-trained to identify facial attributes, bypassing the need for manual training datasets (Schroff et al., 2015). Facenet512

extracts probabilities for ethnicity, gender, age, and emotion⁷. To enhance recognition, we construct a training database of portraits of influential individuals: all Danish ministers and party leaders (2015–2024), Danish royalty, and US presidents, vice presidents, and candidates in the period. Using the alignment component, we crop faces from the portraits, train Facenet512, and identify people in article images out of sample, selecting the highest-probability match per face⁸. This enables us to locate faces, describe attributes, and identify individuals across media. Enforcing alignment prioritises accuracy over efficiency. We extract full probability distributions for attributes (emotions, ethnicities, gender, age) and identification distances, adding approximately 200 hours of processing time across all years. In total, the programme detects 225.254 faces over the ten-year period. Figure 2. illustrates the detection and extracted features⁹.

Text analysis: We represent text information through tokens, analogous to the image analysis, using the multilingual Rapid Automatic Keyword Extraction algorithm (mRAKE) (Piskorski et al., 2021; Rose et al., 2010; Grabovets, 2022), adapted to Danish (see appendix 7.6.). Based on co-occurrence statistics, mRAKE summarises each article in 20 lemmatised keywords, ensuring consistency with image features (Lind, 2021). Full text sentiment scores are estimated using Sentida (Lauridsen et al., 2019a,b; Nimb et al., 2022). Despite limitations in low-resource languages, prior studies validate its use for analysis in Danish (Enevoldsen and Hansen, 2017; Schneidermann and Pedersen, 2022).

3.2 RESULTING DATASET

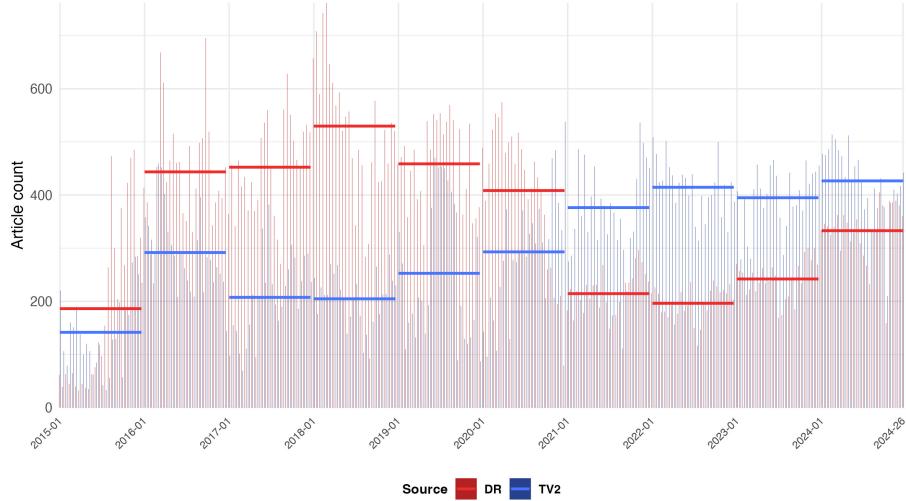


Figure 3: Number of articles per biweek, by source with year average line

Following the procedure above, we obtain tokens reflecting the contents of both the article text and the image accompanying it. We combine tokens from the different extraction

⁷External validity of Facenet512 is measured at 98.4 pct., cf. Serengil (2024), and in combination with RetinaFace surpassing human accuracy (Serengil and Ozpinar, 2024).

⁸See appendix 7.5. for the list of people we recognise.

⁹Facial attributes enable analysis on biases in coverage, presented in Mølby and Bremholm (2025b).

methods and generate a dataset of 168.197 articles \times 140 token columns. We present a descriptive summary of the dataset in table 1., and note that the data is balanced between the two outlets, cf. figure 3. Next, we impose a set of constraints on the token sample, outlined in section 4.5. The process returns on average 2.568 valid tokens per biweek¹⁰.

Table 1: Descriptive summary of image and text tokens

	DR	TV2	Total
Total tokens	6.350.867	5.603.500	11.954.367
Unique tokens	901.517	737.336	1.369.352
Image tag tokens	1.208.652	1.067.478	2.276.130
Unique image tag tokens	45.213	40.131	62.319
Image tag token as pct. of total	19,03%	19,05%	19,04%
Avg. tokens per article	70,50	71,74	71,07
Avg. image tokens per article	13,42	13,67	13,53
Total tokens post filter	-	-	670.191
Total articles	90.088	78.109	168.197
Articles with image tag tokens	85.887	75.657	161.544
Faces recognized	131.240	94.014	225.254
People recognized	15.368	13.240	28.608
People recognized pct.	11,71%	14,08%	12,70%
Faces per article image	1,46	1,20	1,34

3.3 SENTIMENT SCORES OF ARTICLES

Now, we turn to the sentiment score, quantifying how positive or negative a text is¹¹. It provides a simple, initial indication of whether the media differ in news coverage. We take a broad view, examining overall sentiment in DR and TV2 articles from 2015 to 2024.

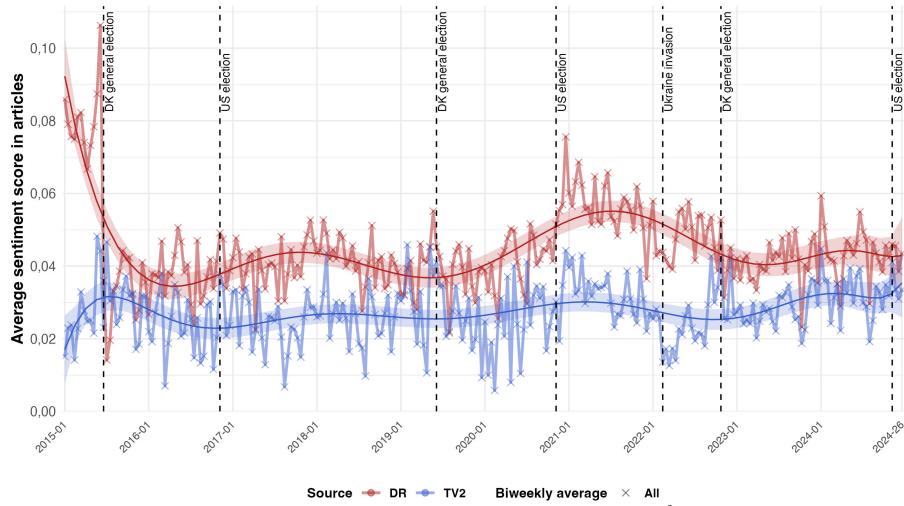


Figure 4: Sentiment score by media, including 10th degree polynomial

The sentiment analysis reveals that DR articles are almost exclusively written in a more positive tone than their TV2 counterparts. In fact, for all but seven biweeks, the

¹⁰Using two-week intervals, the biweekly distribution of tokens is illustrated in appendix 7.7.

¹¹In the following, we use tone, attitude, and sentiment interchangeably, acknowledging that tone and attitude usually have a broader semantic scope than sentiment.

average sentiment of DR articles lies above that of the TV2 articles. Three of the seven two-week periods are in 2015, a year with limited data availability. This is illustrated in figure 4., where a 10th order polynomial, matching the number of years, is fitted to the data-series. We refrain from commenting on the absolute scale of the sentiment scores, as the range is uninformative without a relevant anchor-point.

There is no clear trend in either of the sentiment time-series shown in figure 4. We observe virtually no overlap of the confidence intervals, except for the last few biweeks, which mechanically have broader intervals, as they are at the end of the time-series. This suggests that average sentiment scores from DR and TV2, when approximated by a 10th degree polynomial, are significantly different. The sentiment in both time-series fluctuates, though it appears more stable for TV2 than for DR. This results in expansions and contractions in the gap between the two polynomials, which exhibits a somewhat cyclical pattern. The timing of the cycles is a subject of special interest, as it could shed light on drivers of differences between DR and TV2. In figure 4., vertical lines mark the time of the general elections in Denmark and the US, and the invasion of Ukraine, which might be associated with the cycles. The contractions in the sentiment gap appear to correlate with the timing of the Danish general elections. This suggests that the sentiment gap follows an electoral cycle. We formally investigate this in section 5.

4 MEASURES OF POLARISATION

The cornerstone of this paper is to explore quantitatively whether TV2 and DR present news through systematically different narrative angles in text and lead images, thereby assessing the validity of the working hypothesis: That there is polarisation in Danish media. To identify such differences, we employ three estimators of polarisation adapted from studies of party polarisation in U.S. congressional speech (Gentzkow et al., 2019). These estimators distil high-frequency data into simple measures of difference in tokens between outlets. The resulting measure of either estimator represents the posterior probability that an observer with complete information can accurately determine the source of an article given only a single, randomly drawn token.

Notation of the three subsequent estimators aligns to a certain degree. We assign a unique article number, $i \in \{1, 2, \dots, 168,197\}$, where the source of a given article is given by $S(i) \in \{DR, TV2\}$. For each article, the observed outcome is the vector \mathbf{c}_i , with dimensions $1 \times |\mathbf{J}|$, where $|\mathbf{J}|$ represents the number of unique, valid, tokens present across all articles conditioned on being mentioned at time t . The vector \mathbf{J} contains all valid tokens j at time t . \mathbf{c}_i captures the count of all tokens used in article i . The total count of tokens for article i is given by $m_i = \sum_j c_{i,j}$. Hence, the empirical token frequency

from article i is defined as the vector $\hat{\mathbf{q}}_i = \mathbf{c}_i/m_i$. By summing the observed outcome vector \mathbf{c}_i and m_i over source, we obtain $\hat{\mathbf{q}}^S = \sum_{i \in S} \mathbf{c}_i / \sum_{i \in S} m_i$, i.e. the empirical token frequency of source media. From $\hat{\mathbf{q}}^S$ we compute the token-wise probability of appearing in a DR article as opposed to a TV2 article as:

$$\hat{\rho}_j = \frac{\hat{q}_j^{DR}}{\hat{q}_j^{DR} + \hat{q}_j^{TV2}}$$

Stacking $\hat{\rho}_j$ across all tokens, \mathbf{J} , yields the vector $\hat{\boldsymbol{\rho}}$. We refer to this measure as the source probability, here of DR, where $(\hat{\boldsymbol{\rho}} - 1)$ would be the source probability of TV2. Intuitively, we interpret this as the relative popularity of tokens between the two media. It can be viewed as an editorial preference, where the choice of image and language signals a distinct media profile.

All the aforementioned terms vary over time, denoted by a subscript t in the following. Intuitively, when the vectors $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$ are similar, the articles from the two media outlets are close in terms of linguistic and visual expression. If the vectors are far apart, we interpret it as a higher degree of separation between the media.

4.1 MAXIMUM LIKELIHOOD ESTIMATOR

It is straightforward to construct a maximum likelihood estimator (MLE) on the following form:

$$\hat{\pi}_t^{MLE} = \frac{1}{2} \left(\hat{\mathbf{q}}_t^{DR} \cdot \hat{\boldsymbol{\rho}}_t + \hat{\mathbf{q}}_t^{TV2} \cdot (1 - \hat{\boldsymbol{\rho}}_t) \right) \quad (1)$$

In the above, all tokens not used at time t are excluded from \mathbf{J} . This reduces the choice set in the maximum likelihood estimator to all tokens used at time t . For each media, the dot product of the token frequency and the source probability yields the media-wise polarisation. Intuitively, if the source probability vector was filled with 0, 5, for all tokens j , the media-wise contribution to the MLE polarisation would be 0, 5, as the sum of $\hat{q}_{t,j}^{DR}$ over \mathbf{J} is 1. It follows that the sum of the two media contributions, divided by two, yields total polarisation, $\hat{\pi}_t^{MLE}$.

When $\hat{\boldsymbol{\rho}}_t$ has a large dispersion, the MLE polarisation will intuitively be large, as some tokens are primarily used by just one of the outlets. As this estimator is based on the empirical tokens derived from a finite sample of text and images from the two sources - and not the entire language - the estimated source probability might not capture the true source probability. In the case where a token is used only once at time t by DR, the source probability $\hat{\rho}_{t,j}$ for that token will be 1, effectively amplifying its role in the resulting polarisation measure. However, the occurrence of this token in DR might not

be a signal of strong partisanship towards DR, as it could be down to chance that it is not mentioned in TV2. To clarify the dynamics of this type of estimator, we construct a sandbox example, where each media has two articles and each article has 3 tokens:

$$\begin{aligned} \mathbf{DR}_1 &= \begin{bmatrix} b \\ b \\ d \end{bmatrix}, \quad \mathbf{DR}_2 = \begin{bmatrix} b \\ c \\ d \end{bmatrix}, \quad \mathbf{TV2}_1 = \begin{bmatrix} a \\ a \\ e \end{bmatrix}, \quad \mathbf{TV2}_2 = \begin{bmatrix} c \\ c \\ e \end{bmatrix} \\ j &= \{a, b, c, d, e\}, \quad \hat{\mathbf{q}}^S = \sum_{i \in S} \mathbf{c}_i / \sum_{i \in S} m_i \rightarrow \\ \hat{\mathbf{q}}^{DR} &= \begin{bmatrix} 0/6 \\ 3/6 \\ 1/6 \\ 2/6 \\ 0/6 \end{bmatrix}, \quad \hat{\mathbf{q}}^{TV2} = \begin{bmatrix} 2/6 \\ 0/6 \\ 2/6 \\ 0/6 \\ 2/6 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}} = \begin{bmatrix} 0 \\ 1 \\ 1/3 \\ 1 \\ 0 \end{bmatrix}, \quad \hat{\mathbf{q}}^{DR} \cdot \hat{\boldsymbol{\rho}} = \frac{32}{36}, \quad \hat{\mathbf{q}}^{TV2} \cdot (1 - \hat{\boldsymbol{\rho}}) = \frac{30}{36} \\ \hat{\pi}_t^{MLE} &= \frac{1}{2} \left(\frac{32}{36} + \frac{30}{36} \right) = \frac{62}{72} = \frac{31}{36} \end{aligned}$$

In the above example, we estimate a polarisation close to 1. The full calculations are in appendix 7.8. This illustrates that if each media is unique and there is little overlap between them, the polarisation will be large. Further, when one article has the only representation of a specific token, as in case of a in $\mathbf{TV2}_1$, it greatly influences polarisation. This example further illustrates that when i and J increase, the complexity of vector operations scales linearly, thereby increasing the computational load.

In the MLE, the empirical token frequency of source media $\hat{\mathbf{q}}_t^S$, and the source probability $\hat{\boldsymbol{\rho}}_t$ are both estimated on the same population. This type of estimation introduces a finite sample bias problem, as we outline below.

Finite sample bias: The estimator does not account for the baseline probability of using a token, defined as a token's frequency in "normal" language. As a consequence, some tokens may be labelled as highly partisan to one media outlet, when in fact they are common and non-polarising tokens. Here, common refers to tokens that are generally characterised as part of "normal" language use and should, in principle, be equally distributed across media sources over time. If such common tokens appear in only one media in a given period t , the estimator assigns them as highly partisan, despite their limited polarising effect. In the following, we showcase the origin of this finite sample bias theoretically, using Jensen's inequality (Jensen, 1906) and drawing on Gentzkow et al. (2019). As $\hat{\pi}_t^{MLE}$ is a convex function of $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$, Jensen's inequality implies that this estimator has a positive bias in any finite sample. The estimated source probabilities $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$ are used, not the true probabilities \mathbf{q}_t^{DR} and \mathbf{q}_t^{TV2} , therefore we have an upward bias by Jensen's inequality: $\mathbb{E}[\hat{\pi}_t^{MLE}] \geq \pi_t^{true}$, which arises from inserting estimates into a convex function. To decompose the bias, we compare the expected difference

between the estimated value and the true value:

$$\begin{aligned} & \mathbb{E} \left[\frac{1}{2} \left(\hat{\mathbf{q}}_t^{DR} \cdot \hat{\boldsymbol{\rho}}_t + \hat{\mathbf{q}}_t^{TV2} \cdot (1 - \hat{\boldsymbol{\rho}}_t) \right) - \frac{1}{2} \left(\mathbf{q}_t^{DR} \cdot \boldsymbol{\rho}_t + \mathbf{q}_t^{TV2} \cdot (1 - \boldsymbol{\rho}_t) \right) \right] \\ &= \frac{1}{2} (\mathbf{q}_t^{DR} \cdot \mathbb{E}[\hat{\boldsymbol{\rho}}_t - \boldsymbol{\rho}_t] + \text{Cov}(\hat{\mathbf{q}}_t^{DR} - \mathbf{q}_t^{DR}, \hat{\boldsymbol{\rho}}_t - \boldsymbol{\rho}_t) + \\ & \quad \mathbf{q}_t^{TV2} \cdot \mathbb{E}[(1 - \hat{\boldsymbol{\rho}}_t) - (1 - \boldsymbol{\rho}_t)] + \text{Cov}(\hat{\mathbf{q}}_t^{TV2} - \mathbf{q}_t^{TV2}, (1 - \hat{\boldsymbol{\rho}}_t) - (1 - \boldsymbol{\rho}_t))) \end{aligned}$$

Which we reduce to:

$$\mathbb{E}[\hat{\pi}_t^{MLE}] - \pi_t^{true} = \frac{1}{2} ((\mathbf{q}_t^{DR} - \mathbf{q}_t^{TV2}) \cdot \mathbb{E}[\hat{\boldsymbol{\rho}}_t - \boldsymbol{\rho}_t] + \text{Cov}((\hat{\mathbf{q}}_t^{DR} - \mathbf{q}_t^{DR}) - (\hat{\mathbf{q}}_t^{TV2} - \mathbf{q}_t^{TV2}), \hat{\boldsymbol{\rho}}_t - \boldsymbol{\rho}_t)) \quad (2)$$

The first term stems from $\hat{\boldsymbol{\rho}}_t$ being a non-linear transformation of $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$, and not the true $\boldsymbol{\rho}_t$ value. Given that the non-linearity in $\hat{\boldsymbol{\rho}}_t$ is modest, this component of the bias is small in practice.

The second term represents the primary source of concern in our context, as it is the main driver of upward bias. Since $\hat{\boldsymbol{\rho}}_t$ is constructed directly from $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$, the sampling error in the empirical token frequencies - $\hat{\mathbf{q}}_t^{DR}$ and $\hat{\mathbf{q}}_t^{TV2}$ - transfers mechanically to that of $\hat{\boldsymbol{\rho}}_t$. For example, if $\hat{\mathbf{q}}_t^{DR}$ deviates to some extreme values in a given period, the resulting source probabilities in $\hat{\boldsymbol{\rho}}_t$ will become highly weighted towards DR, thereby inflating the polarisation estimate $\hat{\pi}_t^{MLE}$. Notably, even if the true token frequencies are equal - that is, $\mathbf{q}_t^{DR} = \mathbf{q}_t^{TV2}$ - finite-sample bias alone may lead to polarisation estimates above the true null-polarisation, due to estimation noise.

4.2 LEAVE-OUT ESTIMATOR

We address the severe bias of the MLE by utilising a leave-out (LO) specification of polarisation on the following form:

$$\hat{\pi}_t^{LO} = \frac{1}{2} \left(\frac{1}{|\text{DR}_t|} \sum_{i \in \text{DR}_t} \hat{\mathbf{q}}_{i,t} \cdot \hat{\boldsymbol{\rho}}_{-i,t} + \frac{1}{|\text{TV2}_t|} \sum_{i \in \text{TV2}_t} \hat{\mathbf{q}}_{i,t} \cdot (1 - \hat{\boldsymbol{\rho}}_{-i,t}) \right) \quad (3)$$

The key difference compared with the MLE is the token sample on which $\hat{\mathbf{q}}^S$ and $\hat{\boldsymbol{\rho}}$ are constructed. This specification "leaves out" article i when computing the source probability, hence we denote the new source probabilities $\hat{\boldsymbol{\rho}}_{-i}$ in eq. (3). Now the article-wise choice set of possible tokens excludes tokens used in article i . Following the example outlined in section 4.1., a token mentioned only in one article at time t will no longer have an effect on the polarisation, eliminating the amplifying effect it had in the MLE. As a consequence, we construct article-wise token frequency vectors $\hat{\mathbf{q}}_{i,t}$ and source probability vectors $\hat{\boldsymbol{\rho}}_{-i,t}$. We normalise the dot product by the source-specific total number of articles in t to obtain the LO-estimator of polarisation. To show the implied reduction in the

bias first presented in section 4.1., we outline the bias term for the LO-estimator, but exclusively for DR articles' contribution, for simplicity:

$$\mathbb{E} \left[\frac{1}{|\text{DR}_t|} \sum_{i \in \text{DR}_t} (\hat{\mathbf{q}}_{i,t} \cdot \hat{\boldsymbol{\rho}}_{-i,t} - \mathbf{q}_{i,t} \cdot \boldsymbol{\rho}_t) \right] = \frac{1}{|\text{DR}_t|} \sum_{i \in \text{DR}_t} (\mathbf{q}_{i,t} \cdot \mathbb{E}[\hat{\boldsymbol{\rho}}_{-i,t} - \boldsymbol{\rho}_t]) + \text{Cov}(\hat{\mathbf{q}}_{i,t} - \mathbf{q}_{i,t}, \hat{\boldsymbol{\rho}}_{-i,t} - \boldsymbol{\rho}_t)$$

While this term resembles that of the MLE, the distinction between them is substantial. As $\hat{\boldsymbol{\rho}}_{-i,t}$ is estimated on the population excluding article i , the sampling error becomes independent from that of $\hat{\mathbf{q}}_{i,t}$. This effectively breaks the mechanical relation between the sampling errors of the two, eliminating the second term from equation (2):

$$\text{Cov}(\hat{\mathbf{q}}_{i,t} - \mathbf{q}_{i,t}, \hat{\boldsymbol{\rho}}_{-i,t} - \boldsymbol{\rho}_t) \approx 0$$

Thus, the LO-estimator eliminates the primary source of bias present in the MLE and retains only the bias from the non-linear transformation of $\hat{\boldsymbol{\rho}}_{-i,t}$. By extending the bias term to include TV2 and adopting sample average notation, where $\frac{1}{|\text{S}_t|} \sum_{i \in \text{S}_t} \hat{\mathbf{q}}_{i,t} = \bar{\mathbf{q}}_t^S$, with corresponding true values \mathbf{q}_t^{DR} and \mathbf{q}_t^{TV2} , we obtain the total bias term for the LO-estimator:

$$\mathbb{E}[\hat{\pi}_t^{LO}] - \pi_t^{true} = \frac{1}{2} (\mathbf{q}_t^{DR} - \mathbf{q}_t^{TV2}) \cdot \frac{1}{|\text{DR}_t| + |\text{TV2}_t|} \sum_i \mathbb{E}[\hat{\boldsymbol{\rho}}_{-i,t} - \boldsymbol{\rho}_t] \quad (4)$$

As previously mentioned, this term stems from Jensen's inequality and tends to be small. Above, the notation of estimated source probability $\hat{\boldsymbol{\rho}}_{-i,t}$ differs from that of the true value $\boldsymbol{\rho}_t$. This is because the true (unknown) value of the source probability is fixed at time t , and thus unaffected by dropping one observation.

To illustrate the calculations of polarisation using the LO-specification, we draw on the sandbox example setup from section 4.1. We present the full calculations in appendix 7.8. Applying the framework of the LO-estimator to the same setup yields the following results:

$$\begin{aligned} \hat{\mathbf{q}}_i = \mathbf{c}_i/m_i \rightarrow \quad \hat{\mathbf{q}}_{DR_1} &= \begin{bmatrix} 0/3 \\ 2/3 \\ 0/3 \\ 1/3 \\ 0/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{DR_2} = \begin{bmatrix} 0/3 \\ 1/3 \\ 1/3 \\ 1/3 \\ 0/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{TV2_1} = \begin{bmatrix} 2/3 \\ 0/3 \\ 0/3 \\ 0/3 \\ 1/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{TV2_2} = \begin{bmatrix} 0/3 \\ 0/3 \\ 2/3 \\ 0/3 \\ 1/3 \end{bmatrix} \\ \hat{\boldsymbol{\rho}}_{-DR_1} &= \begin{bmatrix} 0 \\ 1 \\ 1/2 \\ 1 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_{-DR_2} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_{-TV2_1} = \begin{bmatrix} 0 \\ 1 \\ 1/5 \\ 1 \\ 0 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_{-TV2_2} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} \\ \hat{\mathbf{q}}_{DR_1} \cdot \hat{\boldsymbol{\rho}}_{-DR_1} &= 1, \quad \hat{\mathbf{q}}_{DR_2} \cdot \hat{\boldsymbol{\rho}}_{-DR_2} = \frac{2}{3}, \quad \hat{\mathbf{q}}_{TV2_1} \cdot (1 - \hat{\boldsymbol{\rho}}_{-TV2_1}) = 1, \quad \hat{\mathbf{q}}_{TV2_2} \cdot (1 - \hat{\boldsymbol{\rho}}_{-TV2_2}) = \frac{1}{3}, \\ \hat{\pi}^{LO} &= \frac{1}{2} \left(\frac{1}{2} \left(1 + \frac{2}{3} \right) + \frac{1}{2} \left(1 + \frac{1}{3} \right) \right) = \frac{3}{4} \end{aligned}$$

The example demonstrates how the estimation of polarisation differs between the two methods. In this high-polarisation scenario, where the two outlets only share one of five tokens, the LO-estimate yields a lower polarisation than the MLE: $\hat{\pi}^{LO} = \frac{3}{4} < \hat{\pi}^{MLE} = \frac{31}{36}$. The example also serves to illustrate the added complexity of the LO-estimator. Specifically, for each article i , we compute a separate token frequency vector and a corresponding source probability vector $\hat{\rho}_{t,-i}$. As a result, increasing \mathbf{J} by one token, increases the length of $\hat{\mathbf{q}}_{i,t}^{TV^2}$, $\hat{\mathbf{q}}_{i,t}^{DR}$, and $\hat{\rho}_{-i,t}$ vectors. The combined effect of increasing both i and \mathbf{J} leads to an exponential growth in computational complexity. These additional steps impose a substantially greater computational burden in estimating polarisation compared to the MLE.

In consideration of the volume of data outlined in section 3., the computational complexity becomes a limiting factor when estimating polarisation. The LO-estimator is a suitable way to uncover the polarisation in the data, but it does not allow for the use of covariates, and it does not yield estimates of the underlying parameters of the model. The latter prevents a direct analysis of the most polarising tokens, but ensures that the LO-estimator is computationally feasible.

4.3 PENALISED ESTIMATOR

The penalised estimator is proposed by Gentzkow et al. (2019), as the optimal way to reduce finite sample bias, while also generating both token-specific parameter estimates and enabling the use of covariates. In its original form, it incorporates speaker characteristics, which in this context would correspond to article or author attributes. These are omitted here, as comprehensive author data across ten years is unavailable. Instead, a simplified version excluding characteristics is outlined theoretically, though not implemented due to scope and time constraints. The penalised estimator yields coefficient estimates that can be used to plug into the following estimator of token frequency:

$$q_{j,t}^{S(i)} = \frac{e^{u_{i,j,t}}}{\sum_l e^{u_{i,l,t}}}, \quad u_{i,j,t} = \alpha_{j,t} + \varphi_{j,t} \mathbf{1}_{i \in DR} \quad (5)$$

Where l denotes the length of tokens in i . The parameter estimates are derived by minimising the following objective function:

$$\sum_j \left\{ \sum_t \sum_i [m_{i,t} \exp(\alpha_{j,t} + \varphi_{j,t} \mathbf{1}_{i \in DR}) - c_{i,j,t} (\alpha_{j,t} + \varphi_{j,t} \mathbf{1}_{i \in DR}) + \psi |\alpha_{j,t}| + \lambda_j |\varphi_{j,t}|] \right\} \quad (6)$$

This yields estimates of $\{\alpha_t, \varphi_t\}$ to be plugged into eq. (5), which returns a vector of token frequencies by source. These vectors can be used in the following to get estimates of the average polarisation, $\bar{\pi}_t$, following the previously defined definition of $\hat{\rho}_t$:

$$\bar{\pi}_t = \frac{1}{|DR_t \cup TV2_t|} \sum_{i \in DR_t \cup TV2_t} \frac{1}{2} (\hat{\mathbf{q}}_t^{DR} \cdot \hat{\boldsymbol{\rho}}_t + \hat{\mathbf{q}}_t^{TV2} \cdot (1 - \hat{\boldsymbol{\rho}}_t)) \quad (7)$$

The minimand in eq.(6) reflects two key decisions, as outlined by Gentzkow et al. (2019). The first key decision, the likelihood of the multinomial logit model is approximated with the likelihood of a Poisson model, where $c_{i,j,t} \sim \text{Pois}(\exp(\mu_{i,t} + u_{i,j,t}))$ and $\hat{\mu}_{i,t} = \log m_{i,t}$. This approach is adopted given that, fixing $\hat{\mu}_{i,t}$, the likelihood is separable across tokens, where Gentzkow et al. argue that it would otherwise be infeasible to compute. The second key decision made by Gentzkow et al., is to use a Lasso penalty (L_1) in formulating the estimator, which imposes sparsity, penalising high values as it shrinks coefficients towards zero. The appeal of this method is strengthened by its ability to limit the effect of sampling error and thus the source of the finite sample bias. Gentzkow et al. determine the penalty size λ by regularisation path estimation, where a value large enough to satisfy $\varphi_{j,t} = 0$ is chosen as the starting point. By gradually reducing the size of λ_j , and updating parameter estimates, they choose the optimal λ_j , as the value which minimises a Bayesian Information Criterion (BIC). Gentzkow et al. further implement a minimum penalty ψ , which allows numerical convergence while still treating the covariates flexibly. These steps dramatically increase the computational load, given our volume of data and number of individual tokens in \mathbf{J} , as described in section 3.2.

In the context of our analysis, this approach is purely a theoretical exploration of how to deal with finite sample bias in the most complete way. As the limitations of the LO-estimator are not of concern in our context, the penalised estimator entails only marginal benefits compared to the LO-specification, that are greatly outweighed by the added complexity and computational load it requires. Based on all of the above, we consider the LO-estimator the theoretically optimal method in our context.

4.4 INFERENCE

Inference is based on point-wise confidence intervals obtained by random subsampling without replacement of articles. For each biweek t , we draw $K = 100$ subsamples of size $\tau_{k,t} = \tau_t/10$, with τ_t denoting the total number of articles in biweek t . For the k^{th} subsample, we calculate the subsample estimate of polarisation, $\hat{\pi}_t^k$. The confidence interval around the estimate is then:

$$CI_t \equiv (\hat{\pi}_t - (Q_t^k)_{(90)} / \sqrt{\tau_t}; \hat{\pi}_t - (Q_t^k)_{(11)} / \sqrt{\tau_t}) \quad (8)$$

Where $(Q_t^k)_{(c)}$ is the c^{th} order statistic of $Q_t^k = \sqrt{\tau_{k,t}}(\hat{\pi}_t^k - \frac{1}{K} \sum_{l=1}^K \hat{\pi}_t^l)$. Q_t^k being the normalised deviation of $\hat{\pi}_t^k$ from the subsample average, for each subsample k . Sorting

and inserting the 11th and 90th order statistics of Q_t^k into eq. (8) then provides an 80 pct. confidence interval around each estimate of polarisation (Gentzkow et al., 2019; Politis et al., 2001). This procedure, applied across estimators, provides confidence intervals over time and serves as a check of whether estimates differ from the theoretical null, $\hat{\pi}^{\text{null}} = 0, 5$. However, since the true null, in our context, is the polarisation level under randomised article sources, distributed around $\hat{\pi}^{\text{null}}$, we construct a control polarisation to account for this variance. Specifically, we scramble the media vector, resample article vectors without replacement, and re-estimate polarisation for each t , producing $\hat{\pi}_t^*$ which we denote control polarisation. Subsampling again yields $\hat{\pi}_t^{k*}$ and Q_t^{k*} , which, inserted into eq. (8), provide CI_t^* , the point-wise confidence interval for control polarisation. We draw inference by comparing the confidence interval of the estimate to that of the control polarisation. Beyond benchmarking estimated polarisation, control polarisation also reveals estimator bias: Since media affiliation is randomised, $\hat{\pi}_t^*$ should on average equal $\hat{\pi}^{\text{null}} = 0, 5$, and any systematic deviation from this unequivocally documents inherent bias in an estimator.

4.5 TOKEN FILTERING

When constructing the two estimators, the input tokens are of great importance. As visible in the sandbox examples in section 4.1. and 4.2., when a token occurs only once in one article or only in one media, the DR source probability $\hat{\rho}_j$ becomes either 1 or 0, depending on which media used the token. Further, the computational load depends heavily on the length of the J-vector at each time t . We construct a filtering of all tokens before implementing the estimator. First, we filter tokens based on conditions applicable across the entire time-frame. Second, we filter tokens within the individual periods t . In determining the limits and parameters by which we filter, we factor in several aspects, cf. Caprini (2024). These are outlined in appendix 7.9., along with a robustness-check of the chosen parameters and limits (cf. appendix 7.10.).

4.6 ESTIMATED POLARISATION

We estimate polarisation using both the MLE and the LO-estimator to assess polarisation in Danish media. The comparison highlights the advantages of our preferred estimator: As argued theoretically, the MLE suffers from substantial finite-sample bias, which the LO-estimator mitigates by removing the dominant bias term in eq. (2). Figure 5. presents MLE-estimates from 2015–2024 with point-wise confidence intervals and control polarisation. The values are consistently high, often above 0, 575 and for long periods above 0, 6, while control polarisation exceeds 0, 55 in most periods, confirming the expected upward bias. Comparison with the LO-estimate in figure 6. shows a level difference of 0, 05–0, 07 in control polarisation, indicating bias of at least this magnitude. By contrast, the LO-estimates range between 0, 510 and 0, 575 for most periods, with control polarisation

narrowly fluctuating around 0,5, as expected under random assignment. The average control polarisation for the LO-estimator differs from 0,5 only at the 4th decimal. This comparison demonstrates the finite-sample bias of the MLE and the effectiveness of the LO-estimator in reducing it, supporting the theoretical merits of the LO-estimator as a reliable measure of polarisation. Accordingly, the subsequent analysis relies exclusively on the LO-estimates.

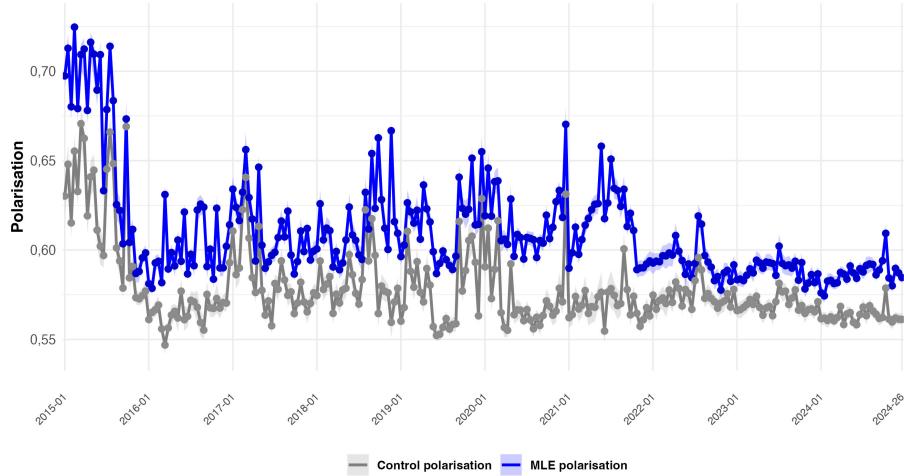


Figure 5: MLE General polarisation 2015-2024

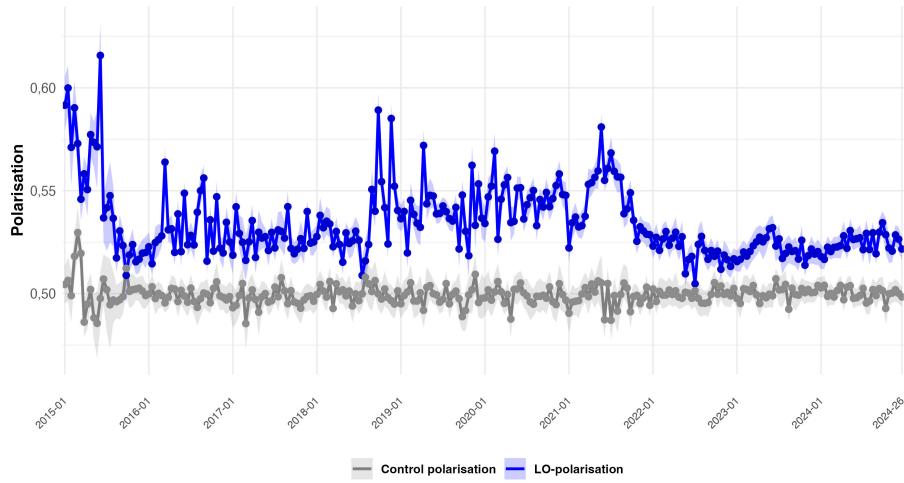


Figure 6: LO General polarisation 2015-2024

In figure 6., we illustrate polarisation from the LO-estimator. The measure is significantly different from control polarisation in all but eight biweeks, with an average level of about 0,534. Polarisation peaks in early 2015 around 0,60. This spike is to a large extend explained by limited data and incomplete section categorisation, with fewer articles the mapping across outlets is unreliable, inflating estimates. From the second half of 2015 through 2018, polarisation stabilises between 0,52 and 0,55, before increasing in late 2018 and remaining relatively high until mid 2021, with estimates frequently above 0,55. Thereafter polarisation declines, and in 2022–2024 it is comparatively low and stable around 0,52–0,53, closely overlapping with U.S. estimates for 2020. Overall,

the evidence points to distinct regimes in polarisation, a pattern that somewhat mimics the development of the sentiment gap between outlets, cf. section 4. The estimated polarisation is robust across various specifications, cf. appendix 7.10. In conclusion, we find significant polarisation between DR and TV2 in our ten-year period, with no sign of a increase over time.

5 ELECTORAL CYCLES IN POLARISATION

Having established significant polarisation in Danish media, we test hypotheses on its dynamics. In [Mølby and Bremholm \(2025b\)](#) we examine correlations across sections, seasonality, conditional variance, regime-changes, and short-term impacts in an event study design, all of which represent novel contributions to the field of media studies. Most prominently, we adapt a framework to analyse fluctuations around electoral cycles below.

The sentiment gap between DR and TV2 fluctuates over time, see figure 7. panel (a). Comparing to figure 7. panel (b) suggests an electoral cycle component in polarisation. Related work documents shifts in tone and coverage intensity before elections, though not in polarisation and not in cyclical patterns ([Hopmann et al., 2012](#)). The observed pattern resembles an “opportunistic” political business cycle, cf. [Schultz \(1995\)](#).

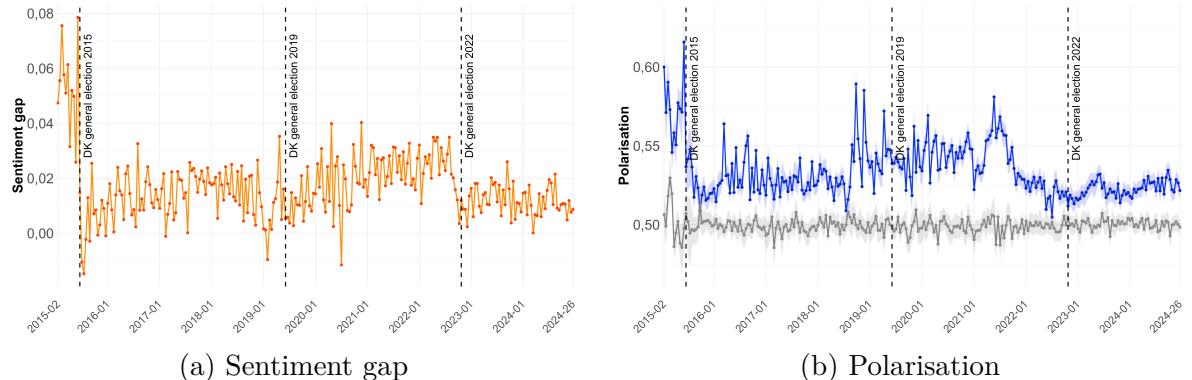


Figure 7: Electoral cycles in the time-series

To test for electoral cycles, we adapt autoregressive models from political business cycle literature, outlined in eq. (9) and (10) ([Davidson et al., 1992](#)). Both include an election-window indicator, $Z_{\text{Election},t}$, capturing level effects in the unconditional mean (eq. 9) and persistence effects through interaction with the AR term (eq. 10). Interaction with the trend component captures deviations from the global trend. The sentiment gap and estimated polarisation follow stationary autoregressive processes with heteroskedasticity¹², and we include quarterly effects to control for seasonality. Reflecting that the time-series includes just three elections, one of which is in 2015, and that token count has no significant effect on polarisation estimates¹², we fit the models to the full time-series including observations from 2015.

¹² See [Mølby and Bremholm \(2025b\)](#) and appendix 7.11. for an analysis of polarisation as a time series.

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 Z_{\text{Election}, t} + [\beta_3 + \beta_4 Z_{\text{Election}, t}] t + \boldsymbol{\gamma}' \boldsymbol{\Omega}_t + \varepsilon_t \quad (9)$$

$$X_t = \beta_0 + [\beta_1 + \beta_2 Z_{\text{Election}, t}] X_{t-1} + [\beta_3 + \beta_4 Z_{\text{Election}, t}] t + \boldsymbol{\gamma}' \boldsymbol{\Omega}_t + \varepsilon_t \quad (10)$$

Parameter estimates of election-cycle effects on the sentiment gap and polarisation are reported in table 2. In models M1–M4 we define an election-window indicator, Z_{election} , equal to 1 in the 13 biweeks (half year) preceding an election. In eq. (9) (M1–M2), the sentiment gap shows a significantly higher unconditional mean prior to elections, in M1, the effect is significant at the 5 pct. level and corresponds to more than a doubling of the unconditional mean¹³. Polarisation shows a smaller, but similarly directed, effect: Its conditional mean is about 3,6 pct. higher in the election window¹⁴. Elevated values in 2015 likely drive these unconditional mean estimates; as data from 2015 is less reliable, we interpret such level effects with caution.

Table 2: Electoral cycles with a leading election window of 13 biweeks, 2015–2024

	Specification			
	M1: Eq (9)	M2: Eq (9)	M3: Eq (10)	M4: Eq (10)
Lag(1)	0,338*** (0,079)	0,453*** (0,070)	0,186** (0,078)	0,455*** (0,069)
Z_{election}	0,016** (0,006)	0,019*** (0,007)		
$\text{Lag}(1) \times Z_{\text{election}}$			0,398*** (0,124)	0,031*** (0,012)
Trend	0,00001 (0,00001)	-0,00001 (0,00001)	0,00001 (0,00001)	-0,00001 (0,00001)
$\text{Trend} \times Z_{\text{election}}$	-0,0001** (0,00004)	-0,0001*** (0,00004)	-0,00004** (0,00002)	-0,0001*** (0,00004)
Constant	0,009*** (0,002)	0,291*** (0,037)	0,012*** (0,002)	0,290*** (0,037)
Quarterly effects	yes	yes	yes	yes
Dependent variable:	Sentiment	Polarisation	Sentiment	Polarisation
Observations	259	259	259	259
R ²	0,333	0,457	0,348	0,454

Note: Robust standard errors in parentheses. *p<0,1; **p<0,05; ***p<0,01

Models M3–M4 estimate eq. (10). Both sentiment and polarisation follow a more persistent autoregressive process before elections, with $\text{Lag}(1) \times Z_{\text{election}}$ positive and significant at the 1 pct. level. None of the specifications indicate a global trend. Within election windows, both series exhibit small but significant negative trends, consistent across M1–M4, and the result holds for various specifications of Z_{election} , cf. appendix 7.13. In summary, both sentiment and polarisation follow electoral cycles, with higher levels and increased persistence prior to elections, followed by convergence as elections

¹³With predictive margins of 0,015 and 0,031 for $Z_{\text{Election}} = (0, 1)$, the sentiment gap is about 106 pct. higher during the election window.

¹⁴Predictive margins for polarisation at 0,533 and 0,552 imply a 3,6 pct. increase.

approach. This indicates a de-polarising effect of elections in Denmark, contrasting with findings from the U.S. ([Fasching et al., 2024](#)).

6 DISCUSSION AND CONCLUSION

Discussion: This paper examines polarisation in Danish media using data from two major outlets. While this narrow scope limits representativeness, the question remains whether the results reflect broader trends. One argument, following [VIVE \(2022\)](#), posits that true polarisation arises mainly from echo-chambers and niche online news outlets, which lie outside our framework. If such outlets drive polarisation, estimates based on mainstream media may underestimate its overall level. Conversely, the competition argument presented in section 2. suggests that low polarisation in DR and TV2 mirrors a generally low demand for polarisation across Danish media. As outlets adjust editorial lines to audience preferences, any wider shift in polarisation would likely affect mainstream media as well. However, this raises an open question as to what governs optimal polarisation.

Conclusion: The overarching objective of the paper has been to test whether there is polarisation in Danish media. We use sentiment analysis as an preliminary offset showcasing differences in tonality between the media. We employ estimators of polarisation based on textual and visual content and find significant polarisation across our ten-year timeline. Based on 11.954.367 tokens from 168.197 articles and images, we estimate polarisation at an average of 0,534, meaning that a token reveals the source outlet with a probability of 53,4 pct. Importantly, polarisation does not increase over time, and we therefore reject the notion of a rising trend in Danish media polarisation. Our analysis combines econometrics with advances in natural language processing and computer vision. Articles and images are processed through keyword extraction, facial analysis, and scene tagging, creating an unprecedentedly rich dataset in the field. We utilise the leave-out estimator, as it has both theoretically and empirically shown a low bias. Robustness checks confirm that polarisation persists across most periods and under stricter thresholds.

Our principal finding - that media-polarisation exists but is not intensifying - corroborates the findings of [VIVE \(2022\)](#) regarding affective polarisation. To the degree that increases in media-polarisation undermine the effective functioning of democratic institutions, our results provide a measure of reassurance. At the same time, the study contributes methodologically by integrating computational techniques with a rigorous econometric framework, enabling quantitative analysis in areas previously inaccessible. We suggest that future research builds on the framework we have outlined in the paper, as it present numerous opportunities to investigate e.g. specific drivers of media polarisation.

7 APPENDIX

7.1 List of R-packages

Package	Link	Package	Link
lubridate	CRAN link	data.table	CRAN link
dplyr	CRAN link	ggplot2	CRAN link
ISOweek	CRAN link	scales	CRAN link
tidyverse	CRAN link	sandwich	CRAN link
lmtest	CRAN link	zoo	CRAN link
purrr	CRAN link	grid	CRAN link
kableExtra	CRAN link	reshape2	CRAN link
gridExtra	CRAN link	stringr	CRAN link
glmnet	CRAN link	xtable	CRAN link
margins	CRAN link	rugarch	CRAN link
forecast	CRAN link	stargazer	CRAN link
tseries	CRAN link	estimatr	CRAN link
forcats	CRAN link	rdrobust	CRAN link

Table 3: R packages used in the analysis with CRAN links

7.2 Competition argument for polarisation

From a micro-economic perspective, media produce news coverage that, if completely objective, would be perfect substitutes of the same product. As economic agents, the media outlets then face incentives to distinguish their coverage from that of competing outlets - Assuming that media sequentially decide both product price and differentiation. They differentiate their products to cater to different consumer tastes by adopting distinct characteristics in their news-products ([Thomas J. Nechyba, 2018](#)). Theoretically, these differences are governed by editorial lines that deviate between media.

7.3 Criteria for relevancy of article links

For both news outlets we aim to discard live updates and news stories in other formats than traditional articles which are accompanied by at least one article image. The web page design is not 1:1 for the two outlets however we discard the same type of unwanted article types by enforcing the following criteria on the URL-string of the article links.

For DR.dk we require that ”/nyheder/” or ”/Nyheder/” is present in the URL and that neither ”/om-dr/”, ”/reel/”, ”/seneste/”, ”/ultra/”, ”/p3/”, ”/tv-guide/”, ”/etik-og-rettelser/”, or ”/det-bedste-fra-dr/” is present.

Likewise, for TV2.dk we discard unrelevant article types by requiring that the URL-string contains either ”nyheder.”, ”nyhederne.”, ”politik.”, ”finans.”, or ”vejret’. The check is case-insensitive. Further, the presence of ”/reel/” or ”/live/” disqualifies the article link.

We post-process by removing duplicates based on articles title, and further filter out any form of ”live” newsfeed, as these articles have distinct titles (e.g. ”livecenter 12:35”) but the exact same contents, and appear several times a day. This post-processing includes removal of COVID-19 update blogs etc.

7.4 Article categories for both media

DR sections	TV2 sections
<i>Indland</i>	<i>Krimi, Trafik, Samfund</i>
<i>Politik</i>	<i>Politik</i>
<i>Udland</i>	<i>Udland</i>
<i>Regionale</i>	<i>Lokalt</i>
<i>Viden, Vejret</i>	<i>Klima, Tech</i>
<i>Penge</i>	<i>Finans, Business, Penge, Erhverv</i>
<i>Detektor</i>	
<i>Kultur</i>	
<i>Webfeature</i>	

We determine the section in which an article is published based on the article category in the URL-string of the article link. The list of sections for both outlets and how they correspond is listed below:

Some sections overlap and e.g. DR.dk articles matching the TV2.dk section *Klima* is found in both *Viden* and *Vejret*. However, the DR.dk section *Viden* overlaps with the TV2.dk section *Tech*. This justifies that we retain a high mapping accuracy in the compiled section *Other*.

7.5 List of people we recognise in images

Name	Party	Name	Party	Name	Party
Alex Vanopslagh	LA	Joy Mogensen	S	Marianne Jelved	RV
Ane Halsboe-Joergensen	S	Kaare Dybvad	S	Margrethe Vestager	RV
Anette Vilhelmsen	SF	Kaare Dybvad Bek	S	Martin Lidegaard	RV
Astrid Krag	S	Kamala Harris	Dem	Mattias Tesfaye	S
Barack Obama	Dem	Karen Ellemann	V	Mette Bock	LA
Benedikte Kiaer	KF	Karen Haekkerup	S	Mette Frederiksen	S
Benny Engelbrecht	S	Karen Jespersen	V	Mette Kierkgaard	M
Bertel Haarder	V	Karsten Lauritzen	V	Mia Wagner	V
Birthe Roenn Hornbech	V	Kirsten Brosboel	S	Mogens Jensen	S
Bjarne Corydon	S	Kristian Jensen	V	Mona Juul	KF
Brian Mikkelsen	KF	Kristian Thulesen Dahl	DF	Morten Boedskov	S
Carina Christensen	KF	Lars Aagaard	M	Morten Dahlin	V
Caroline Stage Olsen	M	Lars Barfoed	KF	Morten Messerschmidt	DF
Christian Friis Bach	RV	Lars Christian Lilleholt	V	Morten Oestergaard	RV
Christian Rabjerg Madsen	S	Lars Loekke Rasmussen	V	Nick Haekkerup	S
Christina Egelund	M	Lea Wermelin	S	Nicolai Wammen	S
Christine Antorini	S	Lene Espersen	KF	Ole Birk Olesen	LA
Claus Hjort Frederiksen	V	Louise Schack Elholm	V	Ole Sohn	SF
Connie Hedegaard	KF	Lykke Friis	V	Pelle Dragsted	Oe
Dan Joergensen	S	Mai Mercado	KF	Per Stig Moeller	KF
Donald Trump	Rep	Mai Villadsen	Oe	Peter Christensen	V
Ellen Trane Noerby	V	Magnus Heunicke	S	Peter Hummelgaard	S
Esbens Lunde Larsen	V	Manu Sareen	RV	Pernille Rosenkrantz-Theil	S
Eva Kjer Hansen	V	Marie Bjerre	V	Pernille Skipper	Oe
Finn Poulsen	KF	Marianne Jelved	RV	Pernille Vermund	NB
Franciska Rosenkilde	Aa	Margrethe Vestager	RV	Pia Kjaersgaard	DF
Gitte Lillelund Bech	V	Martin Lidegaard	RV	Pia Olsen Dyhr	SF
Helle Thorning-Schmidt	S	Mattias Tesfaye	S	Rasmus Helveg Petersen	RV
Henrik Dam Kristensen	S	Mette Bock	LA	Rasmus Jarlov	KF
Henrik Hoeegh	V	Mette Frederiksen	S	Rasmus Prehn	S
Henrik Sass Larsen	S	Mette Kierkgaard	M	Rasmus Stoklund	S
Hillary Clinton	Dem	Mia Wagner	V	Simon Emil Ammitzboell	LA
Ida Auken	SF	Mogens Jensen	S	Simon Kollerup	S
Inger Stoebjerg	V	Mona Juul	KF	Soeren Gade	V
Jacob Jensen	V	Morten Boedskov	S	Soeren Pape Poulsen	KF
Jakob Ellemann-Jensen	V	Morten Dahl	V	Soeren Pind	V
Jakob Engel-Schmidt	M	Morten Messerschmidt	DF	Sofie Carsten Nielsen	RV
Jakob Axel Nielsen	KF	Morten Oestergaard	RV	Sophie Haestorp Andersen	S
J.D. Vance	Rep	Nick Haekkerup	S	Sophie Loehde	V
Jeppe Bruus	S	Nicolai Wammen	S	Stephanie Lose	V
Jeppe Kofod	S	Ole Birk Olesen	LA	Thor Möger Pedersen	SF
Joergen Neergaard Larsen	V	Ole Sohn	SF	Thomas Danielsen	V
Joe Biden	Dem	Pelle Dragsted	Oe	Thyra Frank	LA
Johanne Schmidt-Nielsen	Oe	Per Stig Moeller	KF	Tim Kaine	Dem
Jonas Dahl	SF	Peter Christensen	V	Tim Walz	Dem
Josephine Fock	Aa	Peter Hummelgaard	S	Tina Nedergaard	V
Pernille Rosenkrantz-Theil	S	Pernille Skipper	Oe	Torsten Schack Pedersen	V
Pernille Vermund	NB	Pia Kjaersgaard	DF	Trine Bramsen	S
Pia Olsen Dyhr	SF	Rasmus Helveg Petersen	RV	Troels Lund Poulsen	V
Rasmus Jarlov	KF	Rasmus Prehn	S	Uffe Elbaek	Aa
Rasmus Stoklund	S	Simon Emil Ammitzboell	LA	Ulla Toernaes	V
Simon Kollerup	S	Soeren Gade	V	Villy Soevndal	SF
Queen Margrethe II	-	King Frederik	-	Queen Mary	-

7.6 mRAKE and stopwords in Danish

We use the stop-words presented in table 4. in the mRAKE process. We rely on [Grabovets \(2022\)](#) in our implementation of the mRAKE algorithm.

Table 4: List of Danish stopwords

af	alle	andet	andre	at	begge	da
de	den	denne	der	deres	det	dette
dig	din	dog	du	ej	eller	en
end	ene	eneste	enhver	et	fem	fire
flere	fleste	for	fordi	forrige	fra	få
før	god	han	hans	har	hendes	her
hun	hvad	hvem	hver	hvilken	hvis	hvor
hvordan	hvorfor	hvornår	i	ikke	ind	ingen
intet	jeg	jeres	kan	kom	kommer	lav
lidt	lille	man	mand	mange	med	meget
men	mens	mere	mig	ned	ni	nogen
noget	ny	nyt	nær	næste	næsten	og
op	otte	over	på	se	seks	ses
som	stor	store	syv	ti	til	to
tre	ud	var				

7.7 Biweekly token distribution, 2015-2024

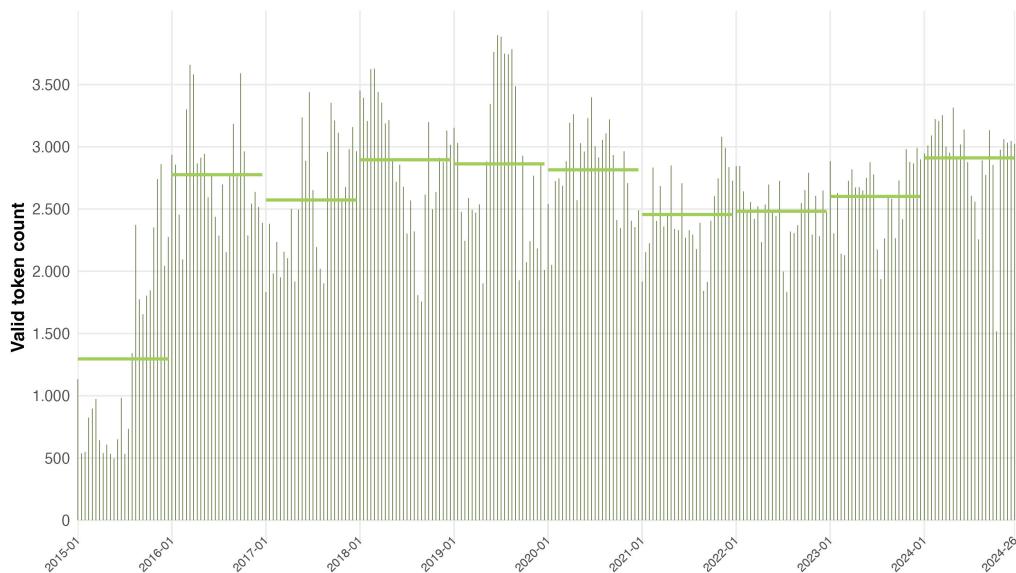


Figure 8: Valid tokens in total

7.8 Full calculations of MLE and LO examples

$$\begin{aligned}
 \mathbf{DR}_1 &= \begin{bmatrix} b \\ b \\ d \end{bmatrix}, \quad \mathbf{DR}_2 = \begin{bmatrix} b \\ c \\ d \end{bmatrix}, \quad \mathbf{TV2}_1 = \begin{bmatrix} a \\ a \\ e \end{bmatrix}, \quad \mathbf{TV2}_2 = \begin{bmatrix} c \\ c \\ e \end{bmatrix} \\
 j &= \{a, b, c, d, e\}, \quad \hat{\mathbf{q}}^S = \sum_{i \in S} \mathbf{c}_i / \sum_{i \in S} m_i \rightarrow \\
 \hat{\mathbf{q}}^{DR} &= \begin{bmatrix} 0/6 \\ 3/6 \\ 1/6 \\ 2/6 \\ 0/6 \end{bmatrix}, \quad \hat{\mathbf{q}}^{TV2} = \begin{bmatrix} 2/6 \\ 0/6 \\ 2/6 \\ 0/6 \\ 2/6 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}} = \begin{bmatrix} 0/(0+2/6)=0 \\ 3/6/(3/6+0)=1 \\ 1/6/(1/6+2/6)=1/3 \\ 2/6/(2/6+0)=1 \\ 0/(0+2/6)=0 \end{bmatrix} \\
 \hat{\mathbf{q}}^{DR} \cdot \hat{\boldsymbol{\rho}} &= \begin{bmatrix} (0/6) \cdot 0=0 \\ (3/6) \cdot 1=3/6 \\ (1/6) \cdot 1/3=1/18 \\ (2/6) \cdot 1=2/6 \\ (0/6) \cdot 0=0 \end{bmatrix} = \left(0 + \frac{3}{6} + \frac{1}{18} + \frac{2}{6} + 0\right) = \frac{32}{36} \\
 \hat{\mathbf{q}}^{TV2} \cdot (1 - \hat{\boldsymbol{\rho}}) &= \begin{bmatrix} (2/6) \cdot 1=2/6 \\ (0/6) \cdot 0=0 \\ (2/6) \cdot 2/3=4/18 \\ (0/6) \cdot 0=0 \\ (2/6) \cdot 1=2/6 \end{bmatrix} = \left(\frac{2}{6} + 0 + \frac{4}{18} + 0 + \frac{2}{6}\right) = \frac{30}{36} \\
 \pi_t^{MLE} &= \frac{1}{2} \left(\frac{32}{36} + \frac{30}{36} \right) = \frac{62}{72} = \frac{31}{36}
 \end{aligned}$$

Leave out sandbox example full calculations:

$$\begin{aligned}
 \hat{\mathbf{q}}_i &= \mathbf{c}_i / m_i \rightarrow \\
 \hat{\mathbf{q}}_{DR_1} &= \begin{bmatrix} 0/3 \\ 2/3 \\ 0/3 \\ 1/3 \\ 0/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{DR_2} = \begin{bmatrix} 0/3 \\ 1/3 \\ 1/3 \\ 0/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{TV2_1} = \begin{bmatrix} 2/3 \\ 0/3 \\ 0/3 \\ 0/3 \\ 1/3 \end{bmatrix}, \quad \hat{\mathbf{q}}_{TV2_2} = \begin{bmatrix} 0/3 \\ 0/3 \\ 2/3 \\ 0/3 \\ 1/3 \end{bmatrix} \\
 \hat{\boldsymbol{\rho}}_{-DR_1} &= \begin{bmatrix} 0/(0+2/6)=0 \\ 1/3/(1/3+0)=1 \\ 1/3/(1/3+2/6)=1/2 \\ 1/3/(1/3+0)=1 \\ 0/(0+2/6)=0 \end{bmatrix}, \quad \hat{\boldsymbol{\rho}}_{-DR_2} = \begin{bmatrix} 0/(0+2/6)=0 \\ 2/3/(2/3+0)=1 \\ 0/(0+2/6)=0 \\ 1/3/(1/3+0)=1 \\ 0/(0+2/6)=0 \end{bmatrix}
 \end{aligned}$$

$$\begin{aligned}
\hat{\rho}_{-TV2_1} &= \begin{bmatrix} 0/(0+0)=0 \\ 3/6/(3/6+0)=1 \\ 1/6/(1/6+2/3)=1/5 \\ 2/6/(2/6+0)=1 \\ 0/(0+1/3)=0 \end{bmatrix}, \quad \hat{\rho}_{-TV2_2} = \begin{bmatrix} 0/(0+2/3)=0 \\ 3/6/(3/6+0)=1 \\ 1/6/(1/6+0)=1 \\ 2/6/(2/6+0)=1 \\ 0/(0+1/3)=0 \end{bmatrix} \\
\hat{\mathbf{q}}_{DR_1} \cdot \hat{\rho}_{-DR_1} &= \begin{bmatrix} 0 \cdot 0 = 0 \\ (2/3) \cdot 1 = 2/3 \\ 0 \cdot 1/2 = 0 \\ (1/3) \cdot 1 = 1/3 \\ 0 \cdot 0 = 0 \end{bmatrix} = 1, \quad \hat{\mathbf{q}}_{DR_2} \cdot \hat{\rho}_{-DR_2} = \begin{bmatrix} 0 \cdot 0 = 0 \\ (1/3) \cdot 1 = 1/3 \\ (1/3) \cdot 0 = 0 \\ (1/3) \cdot 1 = 1/3 \\ 0 \cdot 0 = 0 \end{bmatrix} = \frac{2}{3}, \\
\hat{\mathbf{q}}_{TV2_1} \cdot (1 - \hat{\rho}_{-TV2_1}) &= \begin{bmatrix} (2/3) \cdot 1 = 2/3 \\ 0 \cdot 0 = 0 \\ 0 \cdot 4/5 = 0 \\ 0 \cdot 0 = 0 \\ (1/3) \cdot 1 = 1/3 \end{bmatrix} = 1, \quad \hat{\mathbf{q}}_{TV2_2} \cdot (1 - \hat{\rho}_{-TV2_2}) = \begin{bmatrix} 0 \cdot 1 = 0 \\ 0 \cdot 0 = 0 \\ (2/3) \cdot 0 = 0 \\ 0 \cdot 0 = 0 \\ (1/3) \cdot 1 = 1/3 \end{bmatrix} = \frac{1}{3}, \\
\pi^{LO} &= \frac{1}{2} \left(\frac{1}{|\text{DR}|} \sum_{i \in \text{DR}} \hat{\mathbf{q}}_i \cdot \hat{\rho}_{-i} + \frac{1}{|\text{TV2}|} \sum_{i \in \text{TV2}} \hat{\mathbf{q}}_i \cdot (1 - \hat{\rho}_{-i}) \right) \rightarrow \\
&\pi^{LO} = \frac{1}{2} \left(\frac{1}{2} \left(1 + \frac{2}{3} \right) + \frac{1}{2} \left(1 + \frac{1}{3} \right) \right) = \frac{3}{4}
\end{aligned}$$

7.9 Implementation of token filtering

The most common tokens are likely not informative of the polarisation. This could be tokens like " i ", which is a common token when we split bi- and trigrams into unigrams. In itself, it does not carry any meaning and is likely represented equally between DR and TV2, thus it only contributes to increasing the computational load. Assuming that the most frequent tokens are equally distributed and only contribute to increasing runtime, we choose to filter out the top 0,5 pct. most used tokens. Another aspect is the possible polarisation stemming from very unique articles, e.g. theme weeks concerning very specific topics or other editorial choices that result in highly unique material. To limit the potentially enhancing effect this could have on polarisation, we exclude tokens that are mentioned in less than 10 different periods and tokens that are mentioned less than 50 times across the entire time-frame. These three filters combined constitute the first token filtering step. In the second step we exclude all tokens that are mentioned fewer than three times in the given period. This is done to ensure that when using the LO-estimator for article i containing token j , at least two other articles in t also contain token j . This increases the possibility of getting a meaningful DR probability vector $\hat{\rho}_{-i,j}$. The effect of our filtering is a reduction from 11.954.367 tokens, to 670.191 valid tokens, as presented in table 1. We detail how different criteria-combinations affect estimates of polarisation

in section 7.10.1., as well as outline the impact of varying period-lengths in section 7.10.2.

7.10 Robustness of polarisation estimates

7.10.1 Determining the criteria for the J-vector

In section 4.5., we outline the token filtering process. The second filtering step, excluding all tokens that are mentioned less than three times for period t , remains unadjusted in the following, as it is necessary for the LO-estimator to yield a valid output. The criteria of the first filtering step are not theoretically fixed, therefore the determination of the criteria reflects a "weighted" decision addressing both computational load from an excessive amount of tokens and validity from either too few tokens or selective tokens being filtered out. We test different specifications of the following three criteria in order to find an optimal criteria-combination:

- C.1. Minimum number of times a token is used across all periods.
- C.2. Minimum number of unique periods where a token has to be used at least once.
- C.3. Share of top-frequency tokens excluded based on the total number of times tokens are mentioned across all periods.

To be able to compare the effects of the three criteria, we estimate polarisation using a subset of the possible combinations of the following criteria C.1. $\in \{25, 50\}$, C.2. $\in \{2, 5, 10, 25\}$, and C.3. $\in \{0, 1 \text{ pct.}, 0, 5 \text{ pct.}, 1 \text{ pct.}, 5 \text{ pct.}\}$.

In figure 9., we plot estimates in three years - 2016, 2019, and 2022 - for nine different combinations of the criteria above. We observe that polarisation exists regardless of the criteria combination. We conclude that polarisation is a feature inherent in the tokens themselves and not a reflection of design choices when determining the J-vector.



Figure 9: Specifications of the J-vector and the resulting LO-estimated polarisation for the subset 2016, 2019 and 2022

The specific combination of criteria used to construct our polarisation, is $50 \times 10 \times 0,5$ pct. In figure 9., this iteration is marked as slightly larger pink dots. Compared to the other combinations, we have chosen a conservative level, as to not overstate polarisation. Furthermore, we observe that the underlying fluctuations between the combinations are the same, but the choice affects the level and scale of the estimated polarisation. Another important aspect is that the control polarisation has to be fairly stable around 0,5. This is true for most of the combinations, but the fluctuations vary in scale. We observe that control polarisation for the preferred set of combinations fluctuates less than most of the other sets. Amongst these, a subset of combinations exhibit fairly similar output at the conservative end, and we decide one of these as the preferred combination. Choosing a criteria set that yields conservative estimates is non-controversial and we do not conduct further tests in this regard.

7.10.2 Interval length impact on polarisation

There is no theoretical upper limit to the frequency of estimated polarisation. By defining shorter and shorter time intervals, t , we increase granularity of the estimates. The caveat being that when the intervals shorten, the token input reduces, inferring less reliable estimates. Analysing polarisation on intervals spanning two weeks offers a sensible trade-off between reliability and frequency. We demonstrate this in figure 10., where we plot polarisation in 2024 estimated using weekly, biweekly, and monthly intervals. Mechanically, as intervals increase in length, we observe less volatile estimates with narrower confidence intervals. Confidence intervals are narrow for both biweekly and monthly intervals. When

using weekly intervals the confidence intervals are much wider, and for some weeks overlap with control polarisation, though biweekly and monthly estimates suggest that polarisation is significantly positive. We observe no level differences in the estimated polarisation, again supporting that the LO-estimator yields unbiased estimates even for very limited token input. These aspects motivate that we rely on biweekly estimates of polarisation in the analysis where frequency is particularly critical.

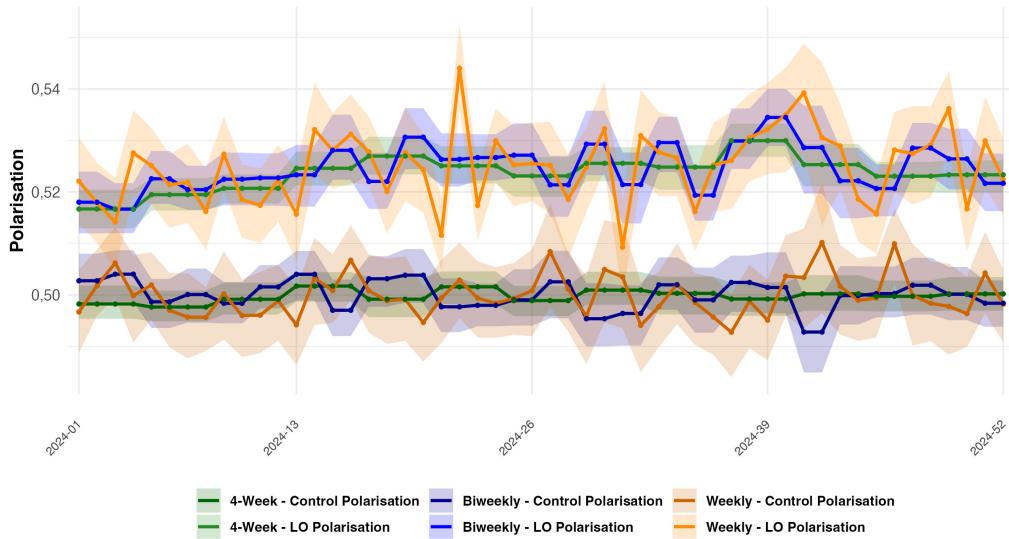


Figure 10: LO-polarisation, variable time interval, 2024

7.10.3 Statistical significance of estimated polarisation

As we explicitly outline in the theoretical foundation for the inference of the polarisation estimate in section 4.4., we employ 80 pct. confidence intervals on all estimates of polarisation throughout the paper. This reflects the practice in the proposal of the method behind the interval (Gentzkow et al., 2019). However, the interval is relatively narrow, and we require stricter confidence on all other estimates in the analysis. Therefore, we examine if polarisation estimates remain significantly different from control polarisation after replacing eq. (8) with eq. (11) below:

$$CI_t \equiv (\hat{\pi}_t - (Q_t^k)_{(98)} / \sqrt{\tau_t}; \hat{\pi}_t - (Q_t^k)_{(3)} / \sqrt{\tau_t}) \quad (11)$$

The equation defines a 95 pct. point-wise confidence interval on polarisation. We illustrate general polarisation using this interval in figure 11. When using an 80 pct. confidence interval, polarisation in 8 biweeks in the period 2015-2024 overlaps with control polarisation, cf. figure 6. This corresponds to 3 pct. of biweeks. In figure 11., the number has increased to 61 biweeks, corresponding to a 20 percentage points increase to 23 pct.

of observations. We conclude that polarisation in some biweeks that were previously significantly different from control polarisation is now insignificant, but notably, for a clear majority of periods, polarisation remains significantly positive. This corroborates the main working hypothesis using a stricter threshold for statistical significance.

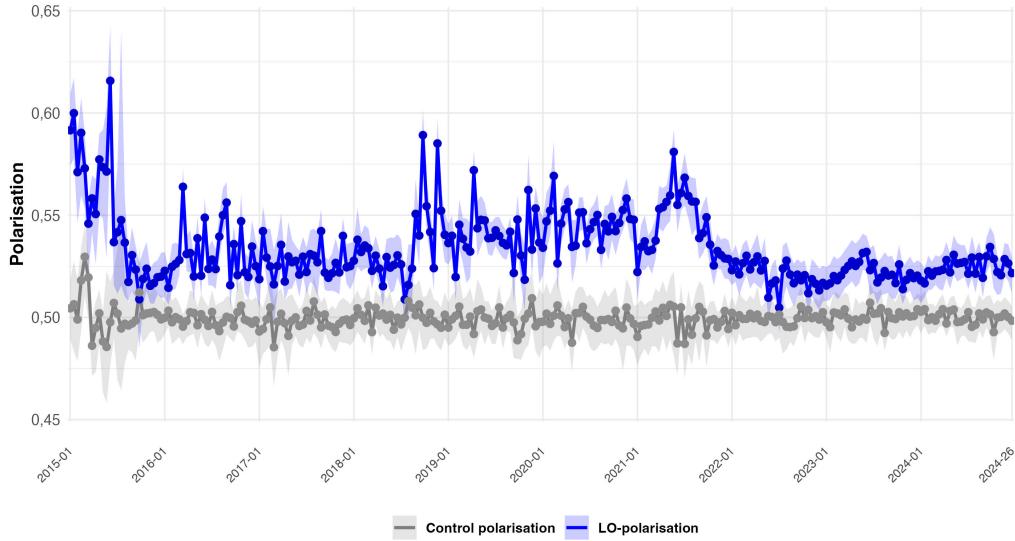


Figure 11: General LO-polarisation, 95 pct. confidence interval

7.11 Characteristics of the polarisation and sentiment gap processes

This note establishes that both polarisation defined by the LO-estimator, $\hat{\pi}_t^{LO}$ and the sentiment gap, $\tilde{X}_t^S = \bar{X}_{DR,t}^S - \bar{X}_{TV2,t}^S$, follow stationary autoregressive processes with heteroskedasticity. If not specified the note draws on test statistics for the period 2015-2024. When the test statistics for the period 2016-2024 deviate from those above we highlight it explicitly. The autoregressive nature of both time-series is presented by the corresponding autocorrelation function plots in figure 12. Further, we report the test statistics for an ADF-test with lag length $k = 1$ in table 5., for both time-series rejecting the null-hypothesis of non-stationarity. We go on to test the range of models used in investigating electoral cycles here exemplified by model M1-M4. For all models we reject the null-hypothesis of homoskedasticity and use HC1 robust std. errors to draw inference from all subsequent estimations, cf. the Breusch-Pagan (BP) statistic in table 5. We do not reject the null hypothesis for homoskedasticity for sentiment gap when looking at 2016-2024, which have no real implications on the results in section 5, cf. table 6. Lastly, based on the Durbin Watson (DW) test statistic we fail to reject the null-hypothesis of no positive autocorrelation in the error term strengthening the case that HC1 robust std. errors are valid for drawing inference based on the selected model specifications.

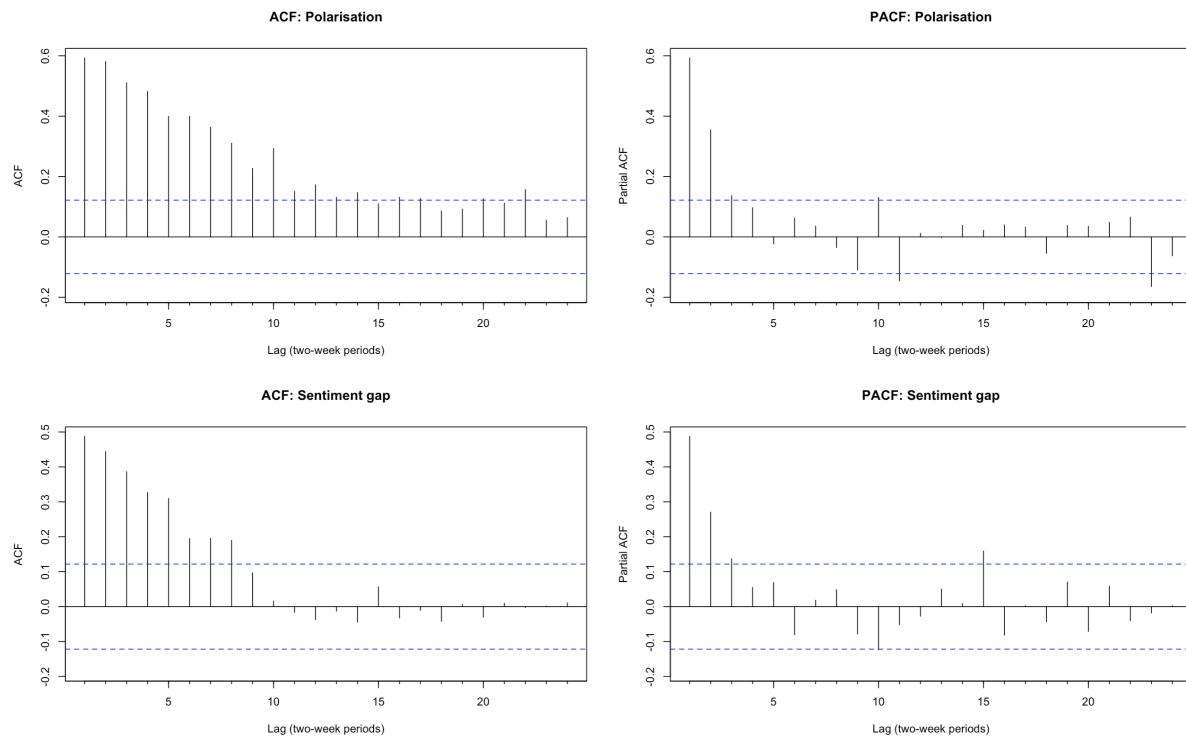


Figure 12: ACF and PACF for polarisation and sentiment gap, 2015–2024.

Table 5: Stationarity, Heteroskedasticity, and Serial Correlation Tests (2015–2024)

Stationarity Tests (Augmented Dickey–Fuller)			
Series	Dickey–Fuller	Lag order	p-value
Polarisation	-5,43	1	0,01
Sentiment gap	-6,49	1	0,01

Breusch–Pagan and Durbin–Watson Tests					
Model	BP stat	df	BP p-value	DW stat	DW p-value
M1: Eq. (9)	52,42	4	$1,13 \times 10^{-10}$	2,14	0,83
M2: Eq. (9)	24,42	4	$6,57 \times 10^{-5}$	2,32	0,99
M3: Eq. (10)	44,37	4	$5,38 \times 10^{-9}$	2,18	0,90
M4: Eq. (10)	25,40	4	$4,18 \times 10^{-5}$	2,33	0,99

Table 6: Stationarity, Heteroskedasticity, and Serial Correlation Tests (2016–2024)

Stationarity Tests (Augmented Dickey–Fuller)			
Series	Dickey–Fuller	Lag order	p-value
Polarisation	-5,19	1	0,01
Sentiment gap	-7,22	1	0,01

Breusch–Pagan and Durbin–Watson Tests					
Model	BP stat	df	BP p-value	DW stat	DW p-value
M1: Eq. (9)	0,96	4	0,92	2,08	0,65
M2: Eq. (9)	10,51	4	0,03	2,35	0,99
M3: Eq. (10)	6,74	4	0,15	2,12	0,78
M4: Eq. (10)	10,52	4	0,03	2,35	0,99

7.12 Front page illustration

We extend our gratitude to Kathrine Andersen Mølby for illustrating the influence of polarisation from online news, cf. figure 13, as used on the front page of our thesis.

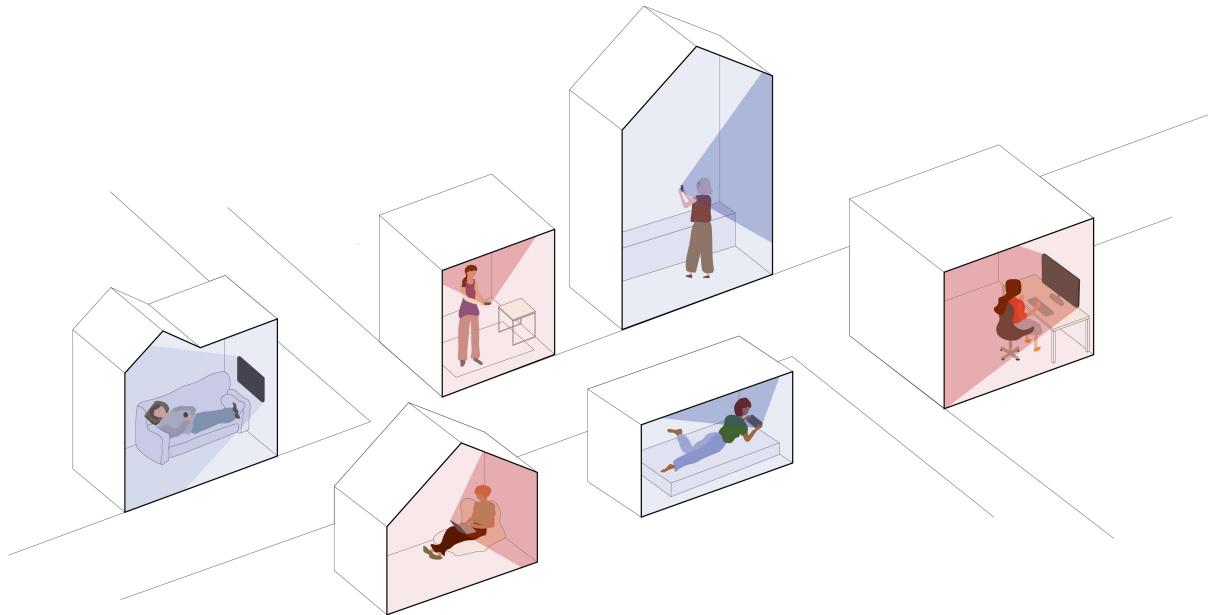


Figure 13: Front page illustration of polarisation in news

7.13 Robust Regression Results: Election window specifications

Table 7: Robust Regression Results: Eq. (9) - Level effects

	Specification:											
	S, l:6	P, l:6	S, l:13	P, l:13	S, l:26	P, l:26	S, c:3	P, c:3	S, c:6	P, c:6	S, c:13	P, c:13
Lag(1)	0,419*** (0,084)	0,538*** (0,070)	0,338*** (0,079)	0,453*** (0,070)	0,365*** (0,075)	0,450*** (0,077)	0,461*** (0,073)	0,558*** (0,069)	0,457*** (0,078)	0,557*** (0,072)	0,449*** (0,073)	0,540*** (0,072)
D _{election}	0,011 (0,010)	0,014 (0,010)	0,016** (0,006)	0,019*** (0,007)	0,012** (0,005)	0,019*** (0,006)	-0,004 (0,010)	0,008 (0,010)	0,002 (0,006)	0,007 (0,006)	0,003 (0,004)	0,006 (0,004)
Trend (biweeks)	-0,000 (0,000)	-0,000* (0,000)	0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,000* (0,000)	-0,000* (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)
Trend × D _{election}	-0,000 (0,000)	-0,000* (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000* (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000*** (0,000)
Constant	0,911*** (0,002)	0,248*** (0,037)	0,248*** (0,002)	0,291*** (0,037)	0,009*** (0,002)	0,291*** (0,041)	0,911*** (0,002)	0,238*** (0,036)	0,010*** (0,002)	0,238*** (0,039)	0,010*** (0,002)	0,246*** (0,038)
Quarterly effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	259 0,302	259 0,434	259 0,333	259 0,457	259 0,325	259 0,462	259 0,299	259 0,424	259 0,290	259 0,427	259 0,292	259 0,431
R ²												

Note: P: Polarisation, S: Sentiment gap. l: Leading window, c: Centred window. The number indicates the number of biweeks prior to the election included. E.g. "P, c:13" refers to estimates for polarisation, using a centred election window with 13 biweeks before and after the election. Standard errors in parentheses are robust. * p<0.1; ** p<0.05; *** p<0.01.

Table 8: Robust Regression Results: Eq. (10) - Dynamic effects

	Dependent variable:											
	S, l:6	P, l:6	S, l:13	P, l:13	S, l:26	P, l:26	S, c:3	P, c:3	S, c:6	P, c:6	S, c:13	P, c:13
Lag(1)	0,454*** (0,079)	0,540*** (0,069)	0,186** (0,078)	0,455*** (0,069)	0,182** (0,084)	0,446*** (0,077)	0,489*** (0,076)	0,559*** (0,067)	0,461*** (0,082)	0,556*** (0,072)	0,204** (0,082)	0,535*** (0,071)
Lag(1) $\times D_{election}$	0,028 (0,180)	0,023 (0,019)	0,398*** (0,124)	0,031*** (0,012)	0,369*** (0,123)	0,032*** (0,011)	-0,142 (0,195)	0,012 (0,018)	-0,003 (0,170)	0,012 (0,012)	0,341*** (0,119)	0,011 (0,008)
Trend (biweeks)	-0,000 (0,000)	-0,000* (0,000)	0,000 (0,000)	-0,000 (0,000)	0,000 (0,000)	-0,000 (0,000)	-0,000* (0,000)	-0,000* (0,000)	-0,000* (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)
Trend $\times D_{election}$	-0,000 (0,000)	-0,000 (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000*** (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000 (0,000)	-0,000*** (0,000)	-0,000*** (0,000)
Constant	0,011*** (0,002)	0,247*** (0,036)	0,012*** (0,002)	0,290*** (0,037)	0,012*** (0,002)	0,295*** (0,041)	0,010*** (0,002)	0,238*** (0,036)	0,011*** (0,002)	0,239*** (0,038)	0,012*** (0,002)	0,249*** (0,038)
Quarterly effects	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Observations	259 0,287	259 0,432	259 0,348	259 0,454	259 0,340	259 0,458	259 0,302	259 0,424	259 0,289	259 0,426	259 0,332	259 0,431

Note: P: Polarisation, S: Sentiment gap. l: Leading window, c: Centred window. The number indicates the number of biweeks prior to the election included. E.g. "P, c:13" refers to estimates for polarisation, using a centred election window with 13 biweeks before and after the election. Standard errors in parentheses are robust. * p<0.1; ** p<0.05; *** p<0.01.

References

- Andris, C., D. Lee, M. J. Hamilton, M. Martino, C. E. Gunning, and J. A. Selden (2015, 04). The rise of partisanship and super-cooperators in the u.s. house of representatives. *PLOS ONE* 10(4), 1–14.
- Ash, E., R. Durante, M. Grebenschikova, and C. Schwarz (2022, April). Visual Representation and Stereotypes in News Media. *CESifo Working Paper Series* (9686).
- Ash, E., S. Galletta, M. Pinna, and C. S. Warshaw (2023). From viewers to voters: Tracing fox news' impact on american democracy. *Journal of Public Economics* 222, 104882. Short communication.
- Berlingske (2010). Er dr-journalister røde? er tv2's blå. Accessed: 2025-05-22. <https://www.berlingske.dk/kronikker/er-dr-journalister-roede-er-tv2s-blaa>.
- Boxell, L., M. Gentzkow, and J. M. Shapiro (2020, January). Cross-country trends in affective polarization. NBER Working Paper 26669, National Bureau of Economic Research, Cambridge, MA. Revised November 2021.
- Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *The American Economic Review* 110(11), pp. 3522–3548.
- Caprini, G. (2023, May). Visual Bias. Economics Series Working Papers 1016, University of Oxford, Department of Economics.
- Caprini, G. (2024, November). Visual bias. Economics series working papers, University of Oxford, Department of Economics.
- Danmarks Radio (2025). Medieudviklingen 2024. Årsrapport, Danmarks Radio. Accessed: 2025-04-12. <https://www.dr.dk/om-dr/fakta-om-dr/dr-analyse/hent-medieudviklingen-2024-som-pdf>.
- Davidson, L. S., M. Fratianni, and J. von Hagen (1992). Testing the satisficing version of the political business cycle: 1905–1984. *Public Choice* 73(1), 21–35.
- Deng, J., J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou (2019). Retinaface: Single-stage dense face localisation in the wild. *CoRR* abs/1905.00641.
- DR (2025). Danmarks radio. Accessed 2025-03-22. <https://www.dr.dk>.
- Enevoldsen, K. C. and L. Hansen (2017). Analysing political biases in danish newspapers using sentiment analysis. *Language Works* 2(2), 88–98.

- Fasching, N., S. Iyengar, Y. Lelkes, and S. J. Westwood (2024). Persistent polarization: The unexpected durability of political animosity around us elections. *Science Advances* 10(36), eadm9198.
- Finlay, P. and C. Argos Translate (2025). Argos translate. GitHub repository: <https://github.com/argosopentech/argos-translate>. Accessed: 2025-02-26.
- Gentzkow, M., J. M. Shapiro, and M. Taddy (2019). Measuring group differences in high-dimensional choices: Method and application to congressional speech. *Econometrica* 87(4), 1307–1340.
- Gift, K. and T. Gift (2015). Does politics influence hiring? evidence from a randomized experiment. *Political Behavior* 37(3), 653–675.
- Grabovets, V. (2022). multi_rake: Multilingual implementation of the rake keyword extraction algorithm. https://github.com/vgrabovets/multi_rake. Accessed: 2025-05-30.
- Hetherington, M. J. and T. J. Rudolph (2015). *Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis*. Chicago: University of Chicago Press.
- Hjorth, F., K. M. Sønderskov, and P. T. Dinesen (2019). Affektiv polarisering i danmark: Et listeeksperiment om social distance til politiske modstandere. *Økonomi & Politik* 92(3). Udgives af Djøf Forlag.
- Hopmann, D. N., P. V. Aelst, and G. Legnante (2012). Political balance in the news: A review of concepts, operationalizations and key findings. *Journalism* 13(2), 240–257.
- Internet Archive (2025). Wayback machine. Accessed: 2025-05-22. <https://archive.org>.
- Iyengar, S., Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science* 22, 129–146.
- Iyengar, S., G. Sood, and Y. Lelkes (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly* 76(3), 405–431.
- Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica* 30, 175–193.
- Journalisten (2010). Danskerne: DR er rød, TV 2 er blå. Accessed: 2025-05-22. <https://journalisten.dk/danskerne-dr-er-rod-tv-2-er-bla/>.

Journalisten (2021). Trafikmålingen lever igen: Her er de 10 største online medier. Accessed: 2025-05-22. <https://journalisten.dk/trafikmaalingen-lever-igen-her-er-de-10-stoerste-online-medier/>.

Kim, E., Y. Lelkes, and J. McCrain (2022). Measuring dynamic media bias. Proceedings of the National Academy of Sciences 119(32), e2202197119.

Kingzette, J., J. N. Druckman, S. Klar, Y. Krupnikov, M. Levendusky, and J. B. Ryan (2021). How affective polarization undermines support for democratic norms. Public Opinion Quarterly 85(2), 663–677.

Lauridsen, G. A., J. A. Dalsgaard, and L. K. B. Svendsen (2019a). Sentida, github repository. <https://github.com/Guscode/Sentida>. Accessed: 2025-05-23.

Lauridsen, G. A., J. A. Dalsgaard, and L. K. B. Svendsen (2019b, Sep.). Sentida paper. Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift 4(1), 38–53.

Lind, S. (2021). Lemmy: A lemmatizer for danish and swedish. <https://github.com/sorenlinde/lemmy>. Accessed: 2025-05-23.

Liu, H. et al. (2024). Llava: Large language and vision assistant. GitHub repository: <https://github.com/haotian-liu/LLaVA>. Accessed: 2025-02-24.

Liu, H., C. Li, Q. Wu, and Y. J. Lee (2023). Visual instruction tuning. In NeurIPS.

Mackinac Center for Public Policy (2023). The overton window. Accessed: 2025-05-26. <https://www.mackinac.org/OvertonWindow>.

Mølby, C. A. and K. A. H. Bremholm (2025a). Feature extraction Github repository. GitHub repository: <https://github.com/Christianmoelby/Feature-extraction-Danish-news>. Contains essential code to scrape and collect Danish news and extract features from text, images and faces. Formats features to a token-dataset and implement polarisation estimators. We reserve the right to continuously update and maintain the repository.

Mølby, C. A. and K. A. H. Bremholm (2025b). Media-Polarisation in Denmark, Beyond the Headlines: Textual and Visual Content from a Decade of News. <https://github.com/Christianmoelby/Feature-extraction-Danish-news/blob/85134c734c8cb0ccbf797305ae440bcfe9dd5337/Media-Polarisation%20in%20Denmark.pdf>. Master's thesis, MSc Economics, University of Copenhagen.

Nimb, S., S. Olsen, and T. Troelsgård (2022). Dds: The danish sentiment lexicon. <https://github.com/dsldk/danish-sentiment-lexicon>. Compiled by Det Danske Sprog- og Litteraturselskab (DSL) and Center for Sprogteknologi (CST), University of Copenhagen. Licensed under CC-BY-SA 4.0 International. Accessed: 2025-05-23.

- Ollama (2024). Llava model on ollama. <https://ollama.com/library/llava>. Accessed: 2025-05-23.
- OpenAI (2021). Clip: Connecting text and images. <https://openai.com/index/clip/>. Accessed: 2025-05-23.
- Piskorski, J., N. Stefanovitch, G. Jacquet, and A. Podavini (2021). Exploring linguistically-lightweight keyword extraction techniques for indexing news articles in a multilingual set-up. In Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation, Online, pp. 35–44. Association for Computational Linguistics.
- Politis, D. N., J. P. Romano, and M. Wolf (2001). On the asymptotic theory of subsampling. Statistica Sinica 11(4), 1105–1124.
- Python Software Foundation (2025). Python language reference, version 3.12. <https://www.python.org/>. Accessed: 2025-05-24.
- R Core Team (2025). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Accessed: 2025-05-22. <https://www.R-project.org/>.
- Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever (2021). Learning transferable visual models from natural language supervision. arXiv preprint arXiv:2103.00020.
- Rose, S., D. Engel, N. Cramer, and W. Cowley (2010). Automatic keyword extraction from individual documents. In M. W. Berry and J. Kogan (Eds.), Text Mining: Applications and Theory, pp. 1–20. John Wiley & Sons.
- Russell, N. J. (2006, January 4). An introduction to the overton window of political possibilities. Mackinac Center for Public Policy. Accessed: 2025-05-26. <https://www.mackinac.org/7504>.
- Schneidermann, N. S. and B. S. Pedersen (2022). Evaluating a new danish sentiment resource: the danish sentiment lexicon, dsl. In Proceedings of the SALLD-2 Workshop @ LREC 2022, Marseille, France, pp. 19–24. European Language Resources Association (ELRA).
- Schroff, F., D. Kalenichenko, and J. Philbin (2015). Facenet: A unified embedding for face recognition and clustering. CoRR abs/1503.03832.
- Schultz, K. A. (1995). The politics of the political business cycle. British Journal of Political Science 25(1), 79–99.

Serengil, S. I. (2024). Deepface: A lightweight face recognition and facial attribute analysis framework. GitHub repository :<https://github.com/serengil/deepface>. Accessed: 2025-02-24.

Serengil, S. I. and A. Ozpinar (2024). A benchmark of facial recognition pipelines and co-usability performances of modules. Bilisim Teknolojileri Dergisi 17(2), 95–107.

Slots- og Kulturstyrelsen (2017). Internettrafik – Mediernes Udvikling i Danmark 2017. Technical report, Slots- og Kulturstyrelsen, København. Accessed: 2025-05-30. https://kum.dk/fileadmin/_mediernesudvikling/2017/Internettrafik_2017.pdf.

The New York Times (2019, February 26). How the politically unthinkable can become mainstream. The New York Times. Accessed: 2025-05-26. <https://www.nytimes.com/2019/02/26/us/politics/overton-window-democrats.html>.

Thomas J. Nechyba (2018). Intermediate Microeconomics - An intuitive approach with calculus. Cengage Learning EMEA. Chapter 26. Product differentiation, subsection A.

TV2 (2025). Tv2 nyheder. Accessed: 2025-05-22 <https://tv2.dk>.

VIVE (2022). Polarisering og tilslutning til demokratiske normer – Danskerne syn på politiske modstandere og demokratiet. København: VIVE – Det Nationale Forsknings- og Analysecenter for Velfærd. Written by Rasmus T. Pedersen, Julian Christensen, and Niels Bjørn Grund Petersen.

Zheng, L., W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica (2023). Judging llm-as-a-judge with mt-bench and chatbot arena.