Investigating the Reciprocal Relationship between News and Parliament

A Study of Strikes in the UK Using Deep-learning Sentiment Analysis and Vector Autoregression

Yente Meijers

Spring 2023

MSc Social Scientific Data Analysis
SIMZ51 – Master Thesis 2023
Faculty of Social Sciences Graduate School
Department of Political Science
Supervisor: Robert Klemmensen

Abstract
Previous research has shown that the relationship between the news and parliament is complex and highly variable. Studies have found that in many countries, parliament has minimal influence on the news. However, in the UK, a reciprocal influence between parliament and the news was found, with the media having a stronger effect. This thesis investigates how the news and parliament affect each other, using the recent strikes in the UK as a case study. The relationship is established through both the frequency of newspaper articles and parliamentary speeches and the sentiment conveyed in these texts. Additionally, the political alignment of Members of Parliament and newspapers is considered in this relationship. This thesis introduces an innovative methodology by using an open-source deep-learning Natural Language Processing model (SiEBERT) for analysing the sentiment in the two text sources, combined with vector autoregression. Vector autoregression models estimate the relationship between the number of articles and speeches and the influence of sentiment and political alignment.

The results indicate that the relationship between the news and parliament is indeed reciprocal, and both have a similar effect size. Importantly, the news reacts much more quickly to parliament than vice versa. Also, the thesis finds that negative news articles and speeches are much more influential than positive ones. The political party of the speaker and the political leaning of the newspaper play an important role in the relationship, with the effects varying across ideologies. This study also concludes that both the majority of news articles and parliamentary speeches are very negative towards the strikes in the UK.

Keywords: parliamentary debate, Hansard, newspaper coverage, UK strikes, agenda setting, BERT, SiEBERT, NLP, deep learning, sentiment analysis, vector autoregression

Word count: 19,789

Acknowledgements

I would like to express my gratitude to my supervisor, Robert Klemmensen and to the director of the very new Social Scientific Data Analysis Master's programme, Christopher Swader. I am also eternally thankful for the support of my friends and my partner, David.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

AIC	Akaike Information Criterion
ADF	Augmented Dickey-Fuller
BERT	Bidirectional Encoder Representations from Transformers
FEVD	Forecast Error Variance Decomposition
LibDem	Liberal Democrats
MP	Member of Parliament
NLP	Natural Language Processing
SiEBERT	Sentiment in English BERT
SNP	Scottish National Party
UK	United Kingdom
VAR	Vector Autoregression

1 Introduction

The relationship between mass media and politics has been the subject of academic research for decades. In particular, the role of newspapers in shaping political discourse and public opinion has been studied extensively. However, much less attention has been given to the reciprocal influence of political institutions, such as parliament, on the media. This thesis aims to explore the dynamic interaction between newspapers and parliament in the context of a specific issue and country: the recent strikes in the United Kingdom. By analysing how these two influence each other, the study seeks to shed light on the complex interplay between media and politics and contribute to a better understanding of the role of the media in the UK. The central research question guiding the thesis is: How do newspapers and parliament influence each other in the context of the strikes in the UK? To address this, the study will employ a combination of deep-learning sentiment analysis and vector autoregression (VAR) to quantitatively analyse newspaper coverage and parliamentary debates. Using innovative methods, this relationship is studied empirically rather than theoretically.

The media and politics are often so intertwined that it can be hard to pinpoint which is influencing which. The news media are closely connected to what is discussed in parliament and as a result, play a part in political decision-making. Because politicians can be influenced by the news, so can their decisions (Garland et al., 2018). In this way, news media are some of the most important players in the political decision field. However, theories of media influence on political decision-making are highly contested and remain inconclusive.

Previous studies have suggested that while the influence can occur in both directions, the dominant side depends on the context (van Aelst & Vliegenthart, 2014). This reciprocal relationship has been studied in different ways, one of which is to study newspaper coverage and parliamentary debates. While in many European-centred studies, the parliament was found to have negligible influence on the media, the results for the UK suggested a reciprocal influence, although the media effect was stronger (van Noije et al., 2008; van Aelst & Vliegenthart, 2014). Therefore, the question remains whether these results in the UK are the consequence of a fluke or if the UK is indeed different from other countries. By introducing more variables such as the political party of the speaker and the political leaning of the newspapers, this thesis seeks to untangle the specifics of the relationship between parliament and news in the UK.

Most of the studies done up to this point have focused on how the relationship plays out for multiple issues rather than studying one topic in depth over time. When it comes to media effects, highly mediatised issues are more likely to influence Members of Parliament because of their prominence (Grossman, 2022, p. 448). An issue such as nationwide rail worker strikes disrupting public transport affects most of the population directly, making it likely to be discussed in the media as well as in politics. Starting in 2021, the UK experienced a surge in the number of strikes in many essential professions, which relates to other current affairs such as the cost-of-living crisis. This is what makes the topic of strikes an excellent case study for investigating the relationship between news and parliament.

Other research looking into the interdependencies of news and parliament used measures of attention, which means that only the number of mentions of a specific issue was used. How political elites discuss the strikes can have an impact on the tone of newspaper coverage, thereby influencing public opinion and support for the strikes. Hence, the tone of debates and articles can play a crucial role in the perceived legitimacy of the strikes. Therefore, instead of issue attention, this thesis uses sentiment as the connecting variable in the VAR models. To do so, this thesis employs SiEBERT, a cutting-edge deep-learning natural language processing (NLP) model, which is open-source and outperforms many other language models (Hartmann et al., 2023).

The thesis is structured as follows. The first section provides an overview of the strikes in the UK, highlighting recent occurrences and previous studies on the media's coverage of strikes. Then, previous research on parliamentary debate transcripts is discussed as well as literature on studying the relationship between the media and politics. Subsequently, the theoretical framework is established to inform the hypotheses, focusing on agenda-setting theory. Next, the methodology lays out the case selection, operationalisation, data collection and data processing. This chapter also explains the use of sentiment analysis and vector autoregression. Following this, the results are presented and analysed, also considering the limitations of the research.

Finally, the thesis concludes by reviewing the major findings and suggesting potential avenues for further research.

2 Industrial Action in the UK

To provide background information on the focal topic, this chapter gives an overview of the recent events related to strikes in the United Kingdom.

2.1 Cost of living and a summer of strikes

In 2021, the UK experienced a surge in inflation and a cost-of-living crisis, resulting in increased prices of essential items such as groceries. Additionally, the energy crisis led to exorbitant utility bills, impacting nearly every individual in the country. This pushed many workers to become increasingly aware of and disgruntled at their pay and working conditions. On top of that, the COVID-19 pandemic has drawn attention to the importance but also terrible working conditions for healthcare staff (Milner, 2022). A large number of essential industries participated in recent strikes, such as firefighters, postal staff, barristers, and healthcare workers.

The first major strike to take place since these crises was in November 2021 with a large Tube strike in London, followed by a three-day strike in December by University staff (Transport for London, 2021). The summer of 2022 saw the largest number of strikes since 30 November 2011 when over a million public sector workers took action over pension reform (Lyddon, 2015). According to data from the Office for National Statistics, the UK lost more working days to industrial action in 2022 than in any year since 1989 (Office for National Statistics, 2023b).

In 2023, inflation does not seem to be stabilising and many strikes are still planned from different sectors. Inflation hit 10.5 percent annually in December 2022, while grocery price inflation hit a record 16.7 percent in January 2023 (Office for National Statistics, 2023a). On 15 March 2023, one of the biggest strike days yet happened with over 400 thousand workers taking up industrial action. This is now thought to be the biggest day of industrial action since the wave of strikes started in late 2021, with teachers, doctors, and tube staff all striking (BBC, 2023).

2.2 A new law

According to Milner, previous waves of strikes in Britain, such as the one during the 1970s, have shown that “industrial action often translates into political action – it’s just a question of when” (2022, p. 43). Until 2023, the most recent law that impacted workers’ right to strike was the 2016 Trade Union Act passed by the Conservative government, which reduced possibilities for legal strike action and impeded worker's freedom to organise and participate in industrial action (Dukes & Kountouris, 2016; Davies & Nophakhun, 2018). This law was strongly opposed by trade unions, as it requires 50 percent of trade union members to vote in a ballot for strike action to be legal, in addition to a 40 percent agreement in the ballot to strike (Trades Union Congress, 2017).

In January of 2023, the Conservative government introduced the Strike (Minimum Service Levels) Bill 2023, which would give the government the power to set minimum service levels in key sectors such as ambulance service and fire rescue during periods of industrial action (Department for Business and Trade, 2023). Immediately the Bill was condemned by the Labour Party and trade union leaders (Walker, 2023). The proposal surprised some Conservative voters, as the 2019 Conservative Party Manifesto only suggested introducing minimum service levels for the transport sector (The Conservative Party, 2019). In early March, the Joint Committee on Human Rights found that the Bill in its state at the time would be "incompatible with human rights law" (JCHR, 2023, n.p.).

The Conservative Party used the disruption caused by the recent strikes as a justification for the legislation. However, the negative portrayal of strikes potentially could have influenced how the Conservative MPs perceived the strikes, focusing on the disruptive consequences rather than the important demands being made by the workers.

2.3 Previous research on strikes

Strikes, especially public sector strikes, affect everyone and are thus an incredibly salient issue. However, little academic research has focused on how newspapers report on strikes nor on how the topic is debated in parliament. Most studies have centred on analysing newspaper coverage of the 1970s period of heightened industrial action in the UK (Lyddon, 2015).

In the United States, a connection between increased newspaper coverage and the occurrence of strikes was found during the 1980s strikes in New York (Erickson & Mitchell, 1996). Flynn (2000), also US-focused, found a positive relationship between the duration of the strike and the media attention when considering 90 large-scale strikes between 1980 and 1991. More recently, a study into how different newspapers framed the teacher strikes in Chicago found that newspapers are on average anti-strike, with 45 percent of all the articles being negative and 20 percent being positive (Gillespie, 2021, p. 10). Using qualitative content analysis, the author was able to determine that most of the negative content was directed at the teacher's union. The union was portrayed very negatively, accused of causing unnecessary drama and having no regard for the impact of their strikes on students and parents. However, this research had a narrow time frame and only focused on one type of strike. One study in New Zealand also found evidence of newspapers portraying a strike negatively, in this case, the nurse strike of 2002-2 (Farrow & O'Brien, 2005).

In the UK, the coverage of other big issues that affect the public, such as economic crises and climate change, are more extensively studied than strikes. Much of the research that did consider newspaper coverage of strikes was done on the increased period of strike activity in the 1970s and 80s or focused on one specific strike rather than on a period of increased strike activity. British newspapers have been considered in studies of other countries such as Radebe's (2006) Master's thesis, studying the coverage of industrial action in South Africa by *The Mail* and *The Guardian*, finding that their attention decreased in the post-apartheid years. O'Neill (2007), focusing on the British Fire Brigades Union Dispute in 2002-3, just like in Gillespie's study in Chicago, found that most newspapers covered the union very negatively and used hostile language to describe the industrial dispute. Investigating the junior doctor strike in England in 2016, Davies and Nophakhun concluded that the UK news media vilify industrial action and that newspapers emphasise the negative impact of strikes instead of the working conditions that the strikes are seeking to alleviate (2018, pp. 111–3). Also on the same topic of the junior doctor strike, Macaulay (2016), using a very small sample, discovered that right-wing newspapers covered the strikes very negatively. Only left-leaning newspapers *The Daily Mirror* and *The Guardian* were found to be in support of the strikes.

Up until now, there has been no comprehensive study of newspaper coverage of the recent strikes in the UK. Most recently in an initial assessment, Milner (2022) claims that the wave of industrial action is a result of imbalances in the economy. Many journalists are comparing it to the 'Winter of Discontent' in the 1970s, which refers to a previous period of heightened strike activity. However, she points out that labour laws resulting from the strikes in the 70s have made it much harder to strike and that this means the current times should be regarded as their own pivotal event (Milner, 2022, p. 41).

When it comes to the specific topics of strikes, the author could find no existing research in the UK that considered how the strikes were discussed in parliament. This highlights the importance of this thesis, not only in its contribution to understanding the relationship between news and parliament but also in how the strikes are discussed in the British parliament.

3 The Relationship between News and Parliament

The relationship between news and parliament is complex and multifaceted, with each having the power to significantly influence the other. The news serves as a conduit through which parliamentary debates and decisions are communicated to the public. On the other hand, parliament can influence the media landscape and utilise the news to advance its own political agenda. This literature review explores how the dynamic relationship between news and parliament has been researched.

3.1 Analysing parliamentary debate transcripts

Parliamentary debates are considered a rich textual source for political scientists, containing a wealth of information about the opinions of Members of Parliament (MPs), policymaking, political parties and the current political climate. This thesis uses computational methods to analyse the transcripts of the House of Commons in the UK, called Hansard. Studies have focused on various aspects of parliamentary debates, such as analysing sentiment, predicting voting behaviour, and identifying patterns of language use. One approach is to use sentiment analysis to understand the emotional content of parliamentary debates, as well as the tone and attitude of individual speakers.

Some of the earliest academics to use machine-learning Natural Language Processing (NLP) methods to analyse parliamentary debate transcripts were Grijzenhout et al. (2010, 2014) in the Netherlands. Combining topic modelling with sentiment analysis, they showed how researchers can use machine learning to analyse a large sample of parliamentary debates.

In the area of parliamentary debate analysis in the UK, Abercrombie and Batista-Navarro’s work is most prominent. Their work builds on parliamentary debate research in other countries, but they were the first to take major steps towards computationally analysing Hansard. The only researchers to attempt this earlier in the UK are Onyimadu et al. (2014), who used a sentiment dictionary to classify sentences in Hansard. The sentiment lexicon did not provide high-quality results, with a 43 percent accuracy compared to human coding. Onyimadu et al. did not do any further work on using computational methods to analyse debate transcripts. Inspired by their work, however, Abercrombie and Batista-Navarro started on a very impressive research streak that improved computational debate text analysis step by step. In combination with manually labelling a subset of speeches, Abercrombie and Batista-Navarro (2018) used two machine-learning-based NLP methods (Support vector machine and Multi-layer perceptron) to analyse the sentiment in Hansard on a speech-level. The results of this research were limited, as they found that using sentiment analysis to predict how MPs voted did not provide useful additional information. Yet, this did not put them off from doing more NLP research. Next, they used a deep-learning language model (Bidirectional Encoder Representations from Transformers, BERT) to identify policy preferences in debate motions (Abercrombie et al., 2019). They found that BERT significantly improved results compared to prior machine-learning methods, which is reaffirmed in their analysis of their corpus for sentiment analysis, 'ParlVote' (Abercrombie & Batista-Navarro, 2020a). Most recently, Abercrombie and Batista-Navarro (2022) have further enhanced their method of analysis with BERT-based models to detect policy stances.

Abercrombie and Batista-Navarro lay a very solid foundation for using NLP to analyse debate texts by showing where the shortcomings and opportunities are while revealing important insights into the debates they analysed. In their paper reviewing the literature on sentiment analysis of parliamentary debates, they note a trend in political science where researchers are increasingly using computational methods to analyse debate transcripts (Abercrombie & Batista-Navarro, 2020b).

Building directly on Abercrombie and Batista-Navarro's work, Sawhney et al. (2021) develop a language model to analyse debate texts alongside political cohesion. Using a BERT-based model to analyse many transcripts, they were able to determine that the Conservative party presents a high political cohesion in the debates.

While these methods are invaluable to improving research methods for analysing parliamentary debate transcripts, the author could find no research to date that has applied these novel NLP methods to investigate the sentiment expressed towards a specific issue. Research on the debate surrounding a particular issue has predominantly used manual methods of content analysis on transcripts. For example, Willis (2017) analysed a corpus of debate transcripts with rudimentary machine-learning methods in combination with manual reading to analyse speeches on climate change.

3.2 Analysing newspaper coverage

The media are increasingly paramount in UK politics. Newspapers are an important source of textual information and can help political scientists research media coverage of political issues (van Noije et al., 2008). Having already reviewed newspaper coverage of strikes, the focus here is on studies using UK newspapers only, as the media landscape varies by country (Wells & Caraher, 2014, p. 1439). In the UK, newspapers have been studied to understand how news can play a role in shaping the debate around political issues. Wells and Caraher (2014) examined newspaper coverage of food banks while Hilton et al. (2014) considered alcohol minimum unit pricing. Both studies found that newspapers in the UK have the power to, directly and indirectly, influence the policy process around mediatised issues. They also note that the ideological leaning of the newspaper plays an important role in how an issue is portrayed.

Besides the qualitative content analysis used in the aforementioned research, some more recent studies used NLP methods to analyse newspaper articles. Shapiro et al. (2022) created a sentiment time series from financial newspapers using sentiment lexicons. Despite being recent, their method does not yield accurate results when compared to more advanced NLP techniques. Rameshbhai and Paulose (2019) used Support Vector Machines to opinion mine newspaper headlines, a machine-learning method that can be less accurate compared to deep-learning models. Sentiment analysis has also been employed to study whether newspapers support a specific political party. In Denmark, Enevoldsen and Hansen (2017) used sentiment analysis to assess political biases in newspapers and were able to find support for this. Researchers in Germany have developed a Sentiment Political Compass to analyse newspaper bias towards political parties (Falck et al., 2019).

3.3 Analysing the relationship between media and parliament

Previous research has suggested that analysing parliamentary debates in combination with media sources provides a very good insight into the political climate of a country. Therefore, combining these two data sources can lead to invaluable insights and many researchers developed ways to join parliamentary debate transcripts and media resources such as newspapers (Juric et al., 2012, 2013; Kleppe et al., 2013). Yet, these works do not extensively analyse the combined datasets and did not come to any conclusions about the relationship between media and politics. Much of the research on the relationship between media and politics has focused on elections (see Walgrave & Van Aelst, 2006 for a helpful overview of different research studying the media's influence on politics) and has found that the media play an important role in politics. Jansen et al. (2019) studied the relationship between the media and parties in the run-up to the 2014 European Parliament Elections by studying newspapers and party press releases. They analysed the relationship using vector autoregression to show that the parties were the main driver during this election period.

However, this thesis focuses on the reciprocal relationship between newspapers and parliamentary debates. In this area, the research is more limited and the findings are more nuanced, as the bidirectionality of the relationship needs to be considered rather than only focusing on how the media impacts politics.

Broadly, this field of research posits that the interplay between parliament and the news is complicated and interdependent, with both influencing each other. As a result, many studies disagree on which is the dominant one. This research area rests on the foundation of agenda-setting theory within the field of mediatisation. According to Garland et al. (2018), UK politics has been mediatised to a large extent, but they base their findings on interviews. Other scholars have more systematically studied the relationship between news and parliament. In a thematic study of Hansard and newspapers, Ellis and Kitzinger (2002) combined the two sources of data to consider the connections between them. Walgrave et al. (2008) and van Noije et al. (2008) were at the forefront of a new wave of studies looking into the relationship between parliament and media. Walgrave et al. (2008) relied on time series to study the media's influence on parliament in Belgium while considering different policy issue types and also found that parliament was significantly affected by newspaper coverage. Van Noije et al. (2008) studied mediatisation by investigating whether the influence of the media agenda on the political agenda has increased. Their study compared the parliament and news in the UK and the Netherlands with structural equation models, concluding that the power of the media over politics had increased. However, in the UK they had a mutual influence on each other, whereas the Dutch news was not controlled by parliament. This research showed important results but focused on election periods, which the researchers note might affect the dynamics of the relationship.

Other research considered this relationship during ‘normal’ times. Most notable in this field are the works of Rens Vliegenthart. Van Aelst and Vliegenthart (2014) coined the term ‘Tango’ to describe the dynamic and reciprocal relationship between parliament and news. They tested their hypotheses using parliamentary questions and newspapers in The Netherlands. References to media in parliamentary questions were used to estimate the reciprocal influence between the media and politics. They found that media coverage leads to parliamentary questions and that those parliamentary questions in turn lead to more media coverage, thus confirming the mutual nature of the relationship. However, this method does not fully allow for the potential influence of parliament on the news within the same model.

In a comparison between the Netherlands and Spain, Vliegenthart and Montes (2014) investigate the relationship between the news and parliament over time using vector autoregression, focusing on the issue of the economic crisis. They find that in Spain, the effect of newspapers on parliament is stronger and more clearly bidirectional than in the Netherlands, which has a multiparty government. In Spain, they also determined that there was a clear partisan difference reflected in the press. In a later study, Vliegenthart et al. (2016) further investigate this relationship in a larger-scale comparative study. The researchers perform general pattern matching between one newspaper for each country and parliamentary questions, accounting for different topics. For estimating the mutual dependency, they used pooled time series models. The conclusion was that whether the media influences parliament or vice versa depends on the issue type and the political system of the country. In all countries, they find that the effect of the media on parliament is present but in some, including the UK, the effect of parliament on the media could also be found. Although the comparative results are valuable and show that overall the media seems to have more of an influence on the parliament than vice versa, it is limited in its method as well as in its case selection. For example, for the UK, only the front page of the Wednesday publication of *The Times* was studied and all material was manually coded. Limiting the news media to one newspaper with a clear political leaning favourable to the ruling Conservative party at the time is unlikely to be a revelatory case for media effects on politics.

Although the research by van Aelst and Vliegenthart (2014) takes into account whether the parliamentary questions come from a governing or opposition party, the studies discussed above do not consider the party of the speaker or the ideological leaning of the newspapers as an important factor. This thesis adds political ideology as a crucial factor to the study of the relationship between news and parliament.

Focusing on a specific issue, Dekker and Scholten (2017) found that media attention for policy issues in the Netherlands leads to them being added to the parliamentary agenda. They use immigration, a highly mediatised issue, and they conclude that not only the quantity but also the tone of the news coverage influenced the policy agenda. Keeping this in mind, this thesis goes beyond using issue attention as a measure of comparing news and parliament but also adds the dimension of sentiment to consider not only *if* an issue was discussed in parliament or the news but also *how*.

Notably, the studies above do not employ NLP methods to analyse the text from either the parliament or the news. Therefore, this thesis develops this area of research by introducing computational techniques. By automating the analysis process and minimising human error, this approach allows for an increase in sample size while also enhancing the robustness and reproducibility of the research. Furthermore, this type of research has not yet been applied to the specific issue of strikes. After having reviewed the important literature, the next chapter will discuss the theoretical foundations of this thesis.

4 Agenda-setting Theory, Framing and Political Ideology

4.1 Agenda setting

Agenda-setting theory informs studies of the relationship between media and politics. In the case of this thesis, that relationship is how the number of articles in newspapers affects the number of speeches in parliament and vice versa.

The original study on agenda setting in the American context by McCombs and Shaw (1972) showed that "the mass media set the agenda for each political campaign, influencing the salience of attitudes toward the political issues" (p. 177). These initial studies of agenda setting focused on communication research and elections, considering how the media influence what people think. In political science, agenda-setting theory is focused specifically on the policy-making process (Baumgartner & Jones, 1993). Combining the communication and political science theories thus looks into the effect of mass media on the political process (Walgrave & van Aelst, 2006). This framework describes and explains how political actors, in this case, Members of Parliament, decide on "their priorities, give attention to or ignore issues, and do, or do not, take decisions or a stance concerning these topics" (Walgrave & van Aelst, 2006, p. 89).

However, the outcomes of the many studies using agenda-setting theory consistently disagree over whether mass media determine the political agenda. This can undoubtedly be attributed to the varying effect of the media across outlets and countries (Walgrave et al., 2008). In the UK, previous research suggests that the effect of the media on parliament significantly increased in 2008 compared to the preceding decade (van Noije et al., 2008).

Other UK-specific studies have found that the reciprocity of the agenda-setting relationship between media and politics is significant, with the political side often having more influence over the media in the UK than in other countries (Vliegenthart et al., 2016) but that the media did affect policy development in certain areas (Garland et al., 2018, p. 496). Therefore, it is important to consider if, for the specific issue of strikes, there is a reason to expect different results.

4.1.1 Typology of issues

According to previous research, the extent to which the media or politics influences each other is dependent on the topic being discussed. (van Aelst & Vliegenthart, 2014; Vliegenthart et al., 2016; Dekker & Scholten, 2017). The most widely used typology of issues by prominent scholars of communication and political science (e.g. McCombs & Valenzuela, 2020) is Soroka's typology (2002, pp. 15–31). Soroka's elaborate typology of issues differentiates between sensational, prominent, and governmental issues. Firstly, sensational issues are often dramatic events which attract attention, leading to a lot of media publicity. Examples of sensational issues are the environment, law and order and transport issues. Next, prominent issues are obtrusive and concrete, which means that they have real-world effects on the public. When it comes to prominent issues, politicians tend to be more reliant on their own opinions which leads the media effects to be weaker than for sensational issues. Examples of prominent issues are economic policy, which individuals experience daily but are still covered by the media (Walgrave et al., 2008, pp. 820–821). Lastly, governmental issues are more abstract and unobtrusive, e.g., government administration, with politicians leading the agenda setting.

Using Soroka's typology of issues, industrial action appears to fall between sensational and prominent. Although the strikes have consequences for everyone, they are not regular enough of a dramatic occurrence to be treated as a simple social or economic policy issue, which is how prominent issues are characterised. Media effects are often the strongest for sensational issues, with prominent issues being more influenced by the politician's observations with more moderate media effects as a result (Walgrave et al., 2008, p. 820).

As the strikes can be considered in between a sensational and a prominent issue as well as previous research suggesting that the media play an important role in strikes, this thesis expects there to be clear effects of the media on politics. Yet, because strikes are not a solely sensational issue, it is not expected that the media will dominate. Therefore, this thesis expects that in the case of strikes in the UK, the relationship between the media and parliament is indeed bidirectional.

H1: Newspaper articles and parliamentary debate speeches have an effect on each other's frequency.

Van Noije et al. (2008) find that when looking at longer times series, the parliament has a delayed reaction to the media, whereas the media reacts more quickly to parliament. The researchers suggest that this instant media reaction at first can make it look like the parliament influences the media more than vice versa. Thus, this thesis also expects to find a difference in reaction time, leading to the following sub-hypothesis:

H1a: The parliamentary debates have a delayed reaction to the news, whereas the news responds quickly to the parliamentary debates.

4.2 Framing

As the quantity and type of media attention go hand-in-hand, it is important to also consider the framing of the strikes. This thesis adds another dimension to the understanding of the relationship between news and parliament by considering sentiment. The study of sentiment expressed in the text sources of this thesis is undergirded by theories of framing.

As opposed to agenda setting, which concerns *whether* the media cover a certain issue, framing is *how* the media present an issue. As argued by McCombs (2013), framing is an extension of agenda setting. A widely cited definition comes from Entman, who wrote that framing is to "select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described" (1993, p. 52). The media can accentuate some features, thus steering the observer towards a certain way of thinking (Chung & Druckman, 2011). Framing is not limited to the media alone, and politicians can also use framing to portray issues in a certain way (Roggeband & Vliegenthart, 2007).

Most commonly, framing is investigated using qualitative content analysis (Cooper et al., 2020). In this thesis, framing is quantitatively analysed by using NLP-based sentiment analysis of the texts to understand if the topic is being discussed negatively or positively. However, no previous research comparing the sentiment expressed in parliament and the news has been done. As a result, there are no firm expectations about the specific relationships that will be found when adding sentiment as an additional variable. Thus, the further hypotheses are loosely defined to guide the quantitative exploration and the rest of the theory section is more suggestive than confirmatory. Parts of the literature point in the broad direction that framing does seem to matter but there is no way to know exactly what to expect because there is no research yet.

Theories on news coverage of strikes posit that newspapers are generally anti-strike (Gillespie, 2021). Other research suggests that the arguments that are effectively presented in the media may contribute to the outcome of industrial action (O'Neill, 2007, p. 813). Therefore, the tone of the newspaper articles is an important factor to consider in studying the news coverage of the strikes. For the debate speeches, the House of Commons is known for its highly emotional debating style. Taking into account that the parties will have clear stances on whether they support or oppose industrial action, the MPs will likely express some type of sentiment towards the strikes (Finlayson, 2017).

Considering the influence of sentiment on political agenda setting, there is previous research that does mention that journalists might be "especially sensitive to negative information" and that news coverage with a "critical, negative, and risk emphasis tends to increase the attention of politicians" (Vliegenthart & Montes, 2014, p. 320). However, Vliegenthart and Montes do not specifically account for this when testing for the relationship. This does suggest that it is likely that sentiment will play a role in how the number of articles in newspapers affects the number of speeches in parliament and vice versa. Therefore, the second hypothesis about this relationship between news and parliament is as follows:

H2: The type of sentiment expressed in a speech or news article influences the relationship between the news and parliament.

Based on the previous research about the effects of negative news, the first sub-hypothesis is as follows:

H2a: The number of negative newspaper articles affects the number of negative and positive parliamentary speeches more than positive articles.

Although there is no directly relevant previous research about negative parliamentary speeches being more influential on the news, this research expects to find that the news will be more likely to report on negative rhetoric. Thus, the second sub-hypothesis reads as follows:

H2b: The number of negative speeches affects the number of negative and positive newspaper articles more than positive speeches.

4.3 Political ideology

Outside of the direct relationship between media and parliament, this thesis expects to find certain differences in the sentiment related to the party membership of the MP and the political ideology of the newspaper.

4.3.1 Party alignment

In the UK, traditionally, the Labour Party is much more connected to the trade unions and the working people and the Conservative Party focuses more on the economic consequences of the strikes (Wells & Caraher, 2014; Davies & Nophakhun, 2018). Therefore, this research also anticipates finding a difference in how different parties in parliament debate the issue of strikes.

According to a YouGov report, trade unions were perceived more negatively after the strikes of the summer of 2022 (Knowles, 2022). This is very much divided along party lines, with 60 percent of Conservative voters believing that trade unions play a negative role and 62 percent of Labour voters perceiving trade unions positively. Additionally, Milner's (2022) brief analysis of the wave of strikes concurs that party responses have been predictable. Conservative politicians exhibit anti-union rhetoric, while Labour has been relatively muted in their response.

Because of the historical alignment of the Labour Party with trade unions and the Conservative Party's opposition to trade unions, hypothesis 3a is as follows:

H3a: Conservative MPs will express more negative sentiment in the speeches about strikes than their Labour and Centre peers.

Besides influencing the sentiment that is expressed, this thesis also expects the political party to play a role in how the number of negative or positive newspaper articles affects the number of negative and positive speeches. Because this is a more exploratory part of this work of research, there are no firm expectations as to how a certain newspaper ideology will affect a certain political party's speeches. Therefore, hypothesis 3b is as follows:

H3b: The influence of distinct newspaper ideologies and their sentiment varies across the speeches of different political parties.

4.3.2 Newspaper ideology

Previous research suggests that the political leaning of newspapers influences how they portray certain issues. Thus, because this thesis considers the specific issue of strikes, it is important to consider the political leaning of the newspapers. Research in Canada found that "a conservative paper– compared to a liberal outlet – preferred to attribute responsibility for the strike and its consequences to the unions" (Knight in Jost & Koehler, 2021, p. 490). Therefore, this research expects newspapers to report according to their political leaning.

Right-wing newspapers such as *The Sun*, *The Daily Mail*, and *The Times* covered the union negatively in the early 2000s (O'Neill, 2007). Only *The Mirror* and *The Guardian* have been found to be in support of strikes (Macaulay, 2016).

More recently, according to several trade union leaders, right-wing newspapers have blamed unions for the effects of strikes such as travel disruptions as a result of the rail workers' strikes. Mick Lynch, the general secretary of the National Union of Rail, Maritime and Transport Workers, has spoken up about this, claiming that right-wing tabloid newspapers such as *The Daily Mail* are "not on the side of the working people" (PoliticsJOE, 2022). This is rooted in the Conservative political side historically not being connected to trade unions, the main conduit through which strikes are organised and the alignment of right-wing newspapers with the Conservative party (Hertel-Fernandez et al., 2021).

In other UK-based research on strikes, *The Sun* and *Daily Mail* have often "taken a consistently antagonistic stance towards attitudes and values which challenge the neoliberal orthodoxy" and have instead expressed "support for increased suppression of trade union activism, including the Trade Union Act 2016" (Davies & Nophakhun, 2018, pp. 110–1). In line with previous research, it is expected that the political leaning of the newspapers will influence how they portray the strikes, and hypothesis 4a is as follows:

H4a: Right-wing newspapers will express more negative sentiment in the articles about strikes than their centre and left-wing counterparts.

The same as with the effect on speeches, the party membership of the MP speaking is expected to influence the different political leanings of the newspapers. Again, this area of the research is more unexplored and therefore there are no firm expectations about how introducing political parties and political ideologies of newspapers will affect how the number of positive or negative speeches influences the number of positive and negative newspaper articles. In Spain, which has a moderately similar political structure to the UK with a single-party government, previous research into the relationship between news and parliament found a partisan difference in the press (Vliegenthart & Montes, 2014). Based on this, hypothesis 4b is as follows:

H4b: The influence of political party alignment of the speeches and their sentiment varies across the articles of different newspaper political ideologies.

5 Methodology

5.1 Case selection and operationalisation

This research uses the recent period of increased strike activity in the United Kingdom to investigate the relationship between the news and parliament as well as the influence of political ideology on the sentiment towards the strikes. As previously outlined, based on the prominent nature of strikes, the case for media effects on politics is likely a revelatory case for investigating this effect (Dekker & Scholten, 2017).

To investigate this interaction between media and politics, this thesis takes parliamentary debate speeches as an operationalisation for politics and newspaper articles for the media. Politics and the media are "heavily intertwined" and are both important participants in the discussion of policy and current affairs (Juric et al., 2013, p. 367). Therefore, combining these two sources in a single analysis can discover important relationships. This approach has also been used before in other research into the relationship between media and politics (e.g. Vliegenthart et al., 2016; Jansen et al., 2019), with the one difference being the use of parliamentary debates instead of parliamentary questions. Parliamentary debates contain information about the positions taken by Members of Parliament (MPs) towards important topics, such as strikes. Debates contain more substantial political information about policy than parliamentary questions and better reflect what is going on in politics. Parliamentary debates capture a deeper discussion of the topic rather than only brief mentions during the questions. The UK parliament has two chambers, the House of Commons and the House of Lords. The House of Commons in the UK is the more important legislative body as well as the target of the most media attention. Therefore, the debate transcripts will be taken from the House of Commons (Abercrombie & Batista-Navarro, 2018).

Newspapers are still considered an important source of political news and they have also adapted to modern audiences by publishing online (Jansen et al., 2019). Newspapers also provide more "in-depth and complete coverage" (Walgrave et al., 2008, p. 818) than other media outlets and are more capable of affecting policymakers. The newspapers included in this thesis were selected because of their high circulation figures, widespread geographical distribution and because they represent different political leanings and diverse readerships. Broadsheet newspapers generally have middle-class readerships, with different newspapers attracting different political leanings. Middle-market newspapers are considered relatively less serious. They tend to appeal to an older, middle-class, right-wing audience. On the other hand, tabloid newspapers are known for being more sensationalist in their approach, with a politically diverse readership that is generally made up of working-class individuals (van Noije et al., 2008).

The central concept of framing is made operative as the tone expressed in the speeches and articles that are about the strikes.

5.2 Ethical considerations

The data included in this thesis does not put any individuals at risk. The debate data is aggregated to a party level. However, the raw transcripts used for this thesis contain mentions of individual MPs. Yet, the debate transcripts are made publicly available by the House of Commons and the MPs are public figures with publicly known political alignments, so the thesis does not introduce any new information about the MPs that could put them at risk. The authors of the newspaper articles are not included in the data. Additionally, the newspapers and parliamentary debate transcripts are published with the intent of being distributed.

5.3 Data

According to Sellers (2009), it is important to study the media and politics together as the influence is likely to be reciprocal, with one potentially stronger than the other. Therefore, this research combines two data sources. The time period for both data sources was the same: 1 November 2021 until 20 February 2023. It was chosen because the increased period of strike activity in the UK started with the London Tube strike in November 2021.

All programming for this thesis was done in Python 3.10.9. All code is available on the author's GitHub (github.com/ymeij/SSDA-Thesis/).

5.3.1 Parliamentary debate speeches

The transcripts of parliamentary debates, known as Hansard, are accessible in XML format through the parliamentary monitoring website theyworkforyou.com. This resource is made available under the Open Parliament license (parliament.uk/siteinformation/copyright) and provides access to all transcripts of debates held in the House of Commons from 1919 to the present.

To obtain the debate transcripts from theyworkforyou, Python was used to interact with ParlParse (github.com/mysociety/parlparse/). ParlParse also provides a dataset that contains the party membership of each MP, which comes from the parliamentary informatics project the Public Whip (publicwhip.org.uk).

5.3.1.1 Data pre-processing

After obtaining the XML files, Python was used to parse the necessary information from the XML file. Only debate topic headings and speech elements were kept, dropping non-speech elements and procedural phrases such as '[laughter]' and 'rose—'. Within the speech elements, the tags for the speaker and the actual text were used to create a dataset that had a date, major and minor topic headings and the speeches with their associated Public Whip ID in it. To get the speaker name and party affiliation, the ID was matched to the one in the people file in the ParlParse GitHub ('people.json'). However, after running data checks, sixteen MPs were missing their party affiliation in the provided file so these were manually added using the information on the Public Whip website. With this additional metadata, the dataset was structured to contain the date, the topic, the party of the speaker and the text of the speech.

After creating the first structured dataset, the topics and speeches were filtered using a keyword search ("strike", "industrial action", "labour dispute", "industrial dispute") in Python. However, because the word 'strike' is ambiguous, an additional filtering step was needed to exclude results that were unrelated to industrial action. A manual subsample of 100 speeches was checked for ones that included a different use of the word 'strike' in common phrases of speech, such as 'strike a balance' or 'strike off' as well as military use, such as 'airstrike'. These phrases and words were added to an anti-keywords list that was then used to override the inclusion of a speech, even if it mentions a keyword. Strict filtering was necessary to optimise the results of the sentiment analysis. By restricting which articles were included, it is more likely that the sentiment score output by the language model is regarding strikes as opposed to other topics.

Table 1 contains information about the final parliamentary debate speeches dataset used for further analysis. As visible in the table, some parties speak a lot more in the House of Commons, which is proportional to the number of MPs. In this thesis, only the top four parties were included: Conservative, Labour, Scottish National Party and Liberal Democrat.

Number of MPs and Speeches per Party

Party	Unique MPs (total MPs in HoC)	Number of Speeches	% of Total Speeches
Conservative	185 (354)	821	54.8%
Labour	117 (197)	509	34.0%
Liberal Democrat	12 (14)	21	1.4%
Scottish National Party	32 (45)	163	10.9%

n = 1,499

Table 1: Number of MPs and speeches per party

5.3.2 Newspapers

The newspaper articles were retrieved from the Nexis newspaper database, using the search terms ‘strike’ OR ‘strikes’ OR ‘industrial action’ OR ‘industrial dispute’ OR ‘industrial disputes’. The search was then further limited to the desired date range (1 November 2021 – 20 February 2023). The additional search criteria were: language set to ‘English’, publication location as ‘UK’, publication type as ‘Newspapers’, and publication name set to the included newspapers (see Table 2 below). This search resulted in 71,117 articles. The articles were downloaded in bulk, with one download set for each day in the selected timeframe to verify that no day was missed in the download cycles.

5.3.2.1 Data pre-processing

After downloading all the newspaper articles, several steps were taken in Python to structure the data. Each article came as a separate Rich Text File and needed to be put into a data frame with metadata prepending the article text content. From the files, the text lines containing the published date, the newspaper in which it was published, the section it was published in, and the body of the text were extracted and subsequently saved as separate columns.

Then, some filtering was necessary to clean the data. First, duplicate articles were dropped if the title, newspaper and published date were all the same. This reduced the sample to around 56,000. The newspaper articles had different section types and to remove references to strikes in different countries as well as different uses of the word 'strike', such as in a sports context, articles from most sections were removed. Subsequently, 192 different section types were identified and manually parsed. Only newspapers with no specified section or ['POLITICS', 'SCOTTISH', 'UK NEWS', 'NEWS ANALYSIS', 'EDUCATION NEWS', 'TRAVEL QUESTIONS', 'NEWS', 'SOCIETY', 'UK', 'HOME NEWS', 'UK POLITICS', 'CAMPAIGNS'] were included, which reduced the dataset to 37,818 articles.

After scanning a subsample, there still appeared to be much irrelevant content in the articles. Thus, a further exclusion criterion was set to filter articles which did not have a specified section or only a page number. Similar to the steps taken to filter the debate speeches, newspaper articles were dropped if they did not contain one of the search keywords ("industrial action", "labour dispute", "industrial dispute", "walk out") in either the title or the body of the article. The word "strike" was not added to this list of keywords because this filtering step was only run on a subsection of the data that contained no section types and was likely to have articles about sports and war. After these steps, the data was reduced to 7,131 articles. Following the manual verification of a subsample, there still appeared to be some irrelevant articles, especially those published in *The Independent*. Thus, an additional anti-keyword filtering was done with only "football" to exclude any articles about sports.

For further analysis, an additional dataset was also created where the newspapers are aggregated by their political ideological leanings.

Political Leaning and Number of Articles per Newspaper

Publication Name	Political Leaning	Number of Articles	% of total
The Guardian	Left-wing	918	14.1%
The Independent	Centre	2,372	36.5%
Metro	Centre	217	3.3%
The Times and Sunday Times	Right-wing	962	14.8%
The Daily Mirror	Left-wing	421	6.5%
The Daily Mail and the Mail on Sunday	Right-wing	541	8.3%
The Daily Telegraph	Right-wing	673	10.3%
The Sun	Right-wing	402	6.2%

Political leaning based on van Noije et al. (2008) and Hughes (2016) *n = 6,506*

Table 2: Political leaning and number of articles per newspaper

5.4 Method

5.4.1 Natural Language Processing

To analyse the large number of debate speeches as well as all of the newspaper articles, this research relies on natural language processing (NLP). NLP is "an interdisciplinary area of research aimed at making machines understand and process human languages" (Kedia & Rasu, 2020, p. 19). Using computational methods for text analysis is less time-consuming and can increase the scope of analysis in terms of the quantity of data as well as the dimensions of analysis (Indulska et al., 2012). The structuring of the data is still done by the researcher and the output is interpreted by them too, which means that computational approaches complement the abilities of the researcher rather than replace them. Furthermore, by using NLP, the researcher remains impartial and without the opportunity to analyse text with a view to support their hypotheses (Matthes & Kohring, 2008).

Because this thesis is interested in how the media and politics discussed the strikes, sentiment analysis will be used to determine the sentiment expressed in the parliamentary debate speeches and the newspaper articles. To perform the sentiment analysis, this thesis will use Sentiment in English Bidirectional Encoder Representations from Transformers (SiEBERT), which is an open-source deep-learning model available on Hugging Face (huggingface.co/siebert/sentiment-roberta-large-english).

5.4.1.1 Deep learning: Transformers

Within NLP, there is a wealth of different language models available that all understand and process text differently. In NLP models, a body of text, often called a corpus, is represented as tokens. A token is a sequence of characters that represents a single unit of meaning. It can be a word, number, punctuation mark, or any other meaningful unit of text that is separated by spaces or punctuation.

Deep-learning models learn representations of data as well as a function to map data representation, and better generalise to unseen texts than other traditional machine-learning approaches. Despite research showing much higher prediction accuracies, deep-learning techniques are not yet a "standard tool for social science researchers that use supervised learning for text analysis" (Wankmüller, 2022, p. 3). This is partially because training a new model can be very time and resource intensive. Fortunately, Hugging Face has an open-source library with many pre-trained NLP models available, which are all available in Hugging Face's Python library called 'Transformers' which is compatible with Python's deep-learning framework PyTorch (Wolf et al., 2020).

The most promising of the deep learning pre-trained models are those that are Transformers-based. Transformers is a type of deep learning that is based on attention mechanisms. Attention mechanisms allow neural networks to interpret text in a more human manner instead of equally interpreting or 'reading' a token in a sentence. The content of a text dynamically influences which parts of the text are focussed on to better interpret a sentence. As a result, Transformers-based models are more capable of analysing texts holistically. The most notable example of a Transformer-based model, which "literally transformed the field of NLP" (Wankmüller, 2022, p. 27), is BERT (Devlin et al., 2019). BERT sought to address concerns around unidirectional relations between text, previous language models could only use preceding information to predict the next word in a sentence and thus ignore how language can be connected and affect the meaning of prior parts of a sentence. In a given text, BERT is trained to fill in the blank, which can be at the start, middle, or end of a text. Texts are analysed holistically not sentence by sentence or word by word. BERT was trained on a large amount of data, consisting of 3.3 billion words (Devlin et al., 2019, p. 4175).

Further improvement of BERT was done by Liu et al.'s (2019) Robustly Optimized BERT Pretraining Approach (RoBERTa). RoBERTa improves upon BERT by changing the pretraining to include more heterogenous data including web corpora (Liu et al., 2019, pp. 5–6). For this thesis, a RoBERTa-based model, specifically SiEBERT, was thus deemed more suitable because it was trained on more diverse data. SiEBERT was finetuned on more than one different type of text source, which improves the generalisability of the model and it, therefore, outperforms several other BERT-based models (Hartmann et al., 2023, p. 78). The model assigns either a positive or a negative label to an instance of text in a corpus. In this thesis, the level of analysis is the parliamentary debate speeches and the newspaper articles, which means that each speech and each article will be assigned a sentiment label.

5.4.1.2 BERT performance

In a recent review of Transformers-based language models in social science text analysis, they all outperform conventional machine-learning algorithms, with RoBERTa slightly outperforming BERT (Wankmüller, 2022). In her tests, Wankmüller (2022) uses the models available from Hugging Face's Transformers library and the performance results are thus almost directly applicable to this piece of research, with the difference being the type of text that was analysed in the tests. Wankmüller analysed pre-existing datasets such as the Wikipedia Toxic Comment Dataset.

What makes BERT-based models even more suitable for this thesis is that they can pick up on domain-specific jargon and are thus by far the best model for analysing parliamentary debate transcripts (Abercrombie et al., 2019; Abercrombie & Batista-Navarro, 2020b; Sawhney et al., 2021). BERT-based models have performed the best compared to other language models in analysing large sets of parliamentary debate transcripts in the UK (Abercrombie & Batista-Navarro, 2020a, 2022).

5.4.2 Data preparation and processing

Data pre-processing in NLP is essential to enhance the quality, consistency, and usability of text data. Currently, the Transformer models on Hugging Face do not allow long text sequences. Processing longer sequences can exceed memory capacity, slow down computation, and reduce the effectiveness of self-attention mechanisms (Wankmüller, 2022). To overcome these limitations, truncation or splitting of long texts into shorter segments is commonly employed. For BERT-related models, the maximum length is 512 tokens so longer speeches and articles were split to prevent automatic truncation. The splitting was done based on the total number of tokens in a document to ensure that the chunks were roughly the same size. Then, the average sentiment of these chunks was taken as the final value to assign the sentiment label to that speech or article. To implement this modified way of feeding text chunks to the language model, Hugging Face's tokeniser available in the 'Transformers' library in Python was adjusted instead of using the default one (Wolf et al., 2020). The source code from the creators of SiEBERT was used as a baseline and PyTorch was used as the deep-learning framework (Hartmann et al., 2023). The tokeniser also took all the regular steps needed to prepare a corpus such as lowercasing, which are important to ensure the model can interpret texts smoothly. Subsequently, the tokenised corpus was fed into the model.

In the debates dataset, there were some days when no speeches mentioned the strikes. Because this causes a discontinuous time series, a value of zero for each party was added to days that had no data because a lack of speeches means that no positive or negative sentiment was expressed towards the strikes. The same process was applied to the newspaper dataset to ensure that a value of zero was added when a specific newspaper did not publish an article on a given date.

Additionally, to address the hypotheses, the debate speeches dataset was filtered to only include the most important parties. Therefore, only the Conservative, Labour, Liberal Democrat and the Scottish National Party were included in further analysis. Because the Liberal Democrats (LibDem) and the Scottish National Party (SNP) both represent the centre, they were merged into one variable (Centre). Accordingly, the newspapers were also aggregated into their ideology types: right-wing, centre, and left-wing.

Then, the two data frames were merged to have the number of positive and negative speeches and articles on each day, split by party and newspaper.

5.4.3 Vector Autoregression

To investigate the relationship between newspaper coverage and parliamentary debates during the increased period of strike activity, this research will employ vector autoregression (VAR) (Brandt & Williams, 2007). VAR analysis estimates the value of each variable in a separate equation based on the variable itself as well as the other variables in the model. This method considers all variables in the model as endogenous, which means that any variable can be affected by another. This allows the anticipated reciprocal effect between the news media and politics (Vliegenthart & Montes, 2014). Thus, VAR is commonly used in agenda-setting research (e.g.van Noije et al., 2008; Barberá et al., 2019; Jansen et al., 2019; Vliegenthart & Damstra, 2019) because it allows the researcher to estimate the potential causal impact of one variable on the other and vice versa and can assess interdependencies between the two because the two variables can influence each other in the same model. For this reason, "each VAR model contains as many regressions as the number of included variables" (Jansen et al., 2019, p. 16).

In VAR models, different lags can be used to explore the different periods it takes for variables to impact each other. As the time series data of this thesis is at a daily level, one lag is equal to one day. The number of lags determines the number of past observations used to predict future values and can have a significant impact on the accuracy and stability of the model. To allow for response time between the publishing of newspaper coverage about a strike, the VAR analysis for this thesis will explore different lags. The output of the VAR model will suggest which lag significantly predicts a variable of interest. Relatively small lags are expected because previous research has shown that the influence of the media on parliament happens in a short time period while still allowing time for the other party to respond (Walgrave et al., 2008).

The VAR model will estimate which variables at which lags appear to contribute to the value of the outcome variable. For example, if taking the number of negative articles published in right-wing newspapers, the VAR model will list all the other variables at all the lags allowed (for example, 14 days) and show which variable and which lag is the one that helps predict the number of negative articles published in right-wing newspapers. Statistically, this means that by adding the values of another variable, the value of the negative articles published is more accurately predicted than by only using the past values of the negative articles.

The base equation for a VAR model is a system of linear equations that describe the dynamic relationship between multiple time series variables. The base equation for a VAR model with more than two variables with p lags can be written as:

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t$$

In this equation, y_t represents a vector of the values of the variables at time t, and y_{t-1}, y_{t-2} … y_{t-p} represent the lagged values of the variables. The coefficients A_1, A_2, ..., A_p represent how the lagged values of the variables affect the current values, and the intercept term c represents a constant value that affects all the variables equally. The error term ε_t represents the random variation in the data that is not explained by the other factors. If there are more variables, this equation would simply expand according to the number of variables included in the model.

After the model has been fitted, it can be determined which predictor variables are significant for the equation of an outcome variable of interest. The selection of the appropriate predictor variable(s) can be determined by taking the p-value provided by the model, which indicates statistical significance, and if the p-value is lower than the desired threshold of 0.05, that variable at that lag is selected. The variable(s) at the specific lag(s) with significant p-values is then used to test for causality between the predictor(s) and the outcome. For VAR models, the most common way to do this is by using Granger causality tests, which are performed on two variables at a time. Granger causality "tests the null hypothesis that the sum of all lags for a given variable in a single equation is equal to zero. These tests provide a snapshot of causal effects; they do not indicate the sign of the effect" (Soroka et al., 2002, p. 274). So, variable x Granger causes variable y if the prediction for y, which is partially based on its past values, significantly improves when adding the past values of x to the prediction equation.

Another important measure to assess the influence of a predictor variable on the outcome variable while still accounting for the other variables in the model is by using forecast error variance decomposition (FEVD). It is used to quantify the contribution of each variable in the model to the forecast error variance of each of the endogenous variables because, in a VAR model, the endogenous variables are assumed to be affected by their past values as well as the past values of other variables in the system. FEVD calculates the portion of the forecast error variance of a variable that can be attributed to each variable in the system. Thus, FEVD values provide information on how much of the variation in each variable's forecast error is accounted for by shocks to other variables in the system (Brandt & Williams, 2007). This information can be used to understand the relative importance of each variable in the system.

To test hypotheses 1 and 1a, the first VAR model uses the number of newspaper articles as the measure for newspaper coverage, leaving out the sentiment label and political ideology of the newspaper. The number of speeches was used as the measure for parliamentary debates, omitting the sentiment and political party of the speaker. The results of this model will also include an indication of the lag at which the news potentially affects the debates and vice versa.

To test hypotheses 2, 2a and 2b, the sentiment labels assigned to the articles and speeches are brought into the next VAR model. To account for hypothesis 3b and hypothesis 4b the political ideology of the newspaper and the political party of the MP are added to VAR model 3.

For all these models, Granger causality tests and FEVD will be used to investigate causal relationships after the variables of interest and their lags have been selected.

6 Results and Analysis

This chapter contains links to interactive versions of all the graphs. Simply hold control on the keyboard and click on the plot to open it in a web browser.

6.1 Descriptive statistics

The data for Hansard can be found on the author's GitHub alongside the code for getting the results. The data for the newspaper articles are not included because they are accessed via Nexis. All the information to replicate the newspaper search results has been provided.

6.1.1 Hansard

As previously described, the Hansard data was analysed on a speech level per party per day. Below, the sentiment results are summarised by party. In Table 3, it is visible that only the Conservative Party speaks positively half of the time, whereas the Labour Party and Centre parties have a negative tone in the large majority of their speeches. There is also a large difference in the total speeches by party, with Conservative MPs speaking more than twice as often as Labour MPs and more than four times as much as Centre MPs.

Negative and Positive Speeches per Party

Party	Positive Speeches	% Positive	Negative Speeches	% Negative	Total Speeches
Conservative	467	56.9%	354	43.1%	821
Labour	110	22.3%	384	77.7%	494
Centre (LibDem + SNP)	55	29.9%	129	70.1%	184

n=1,499

Table 3: Negative and positive speeches per party

As visible in Figure 1 there are two periods where the number of speeches discussing strikes spikes. The first period is during the summer of 2022 when many professions took industrial action at the same time. During this period, both the Conservative and Labour parties have similar negative compared to positive speeches, although the Labour Party has more negative than positive speeches.

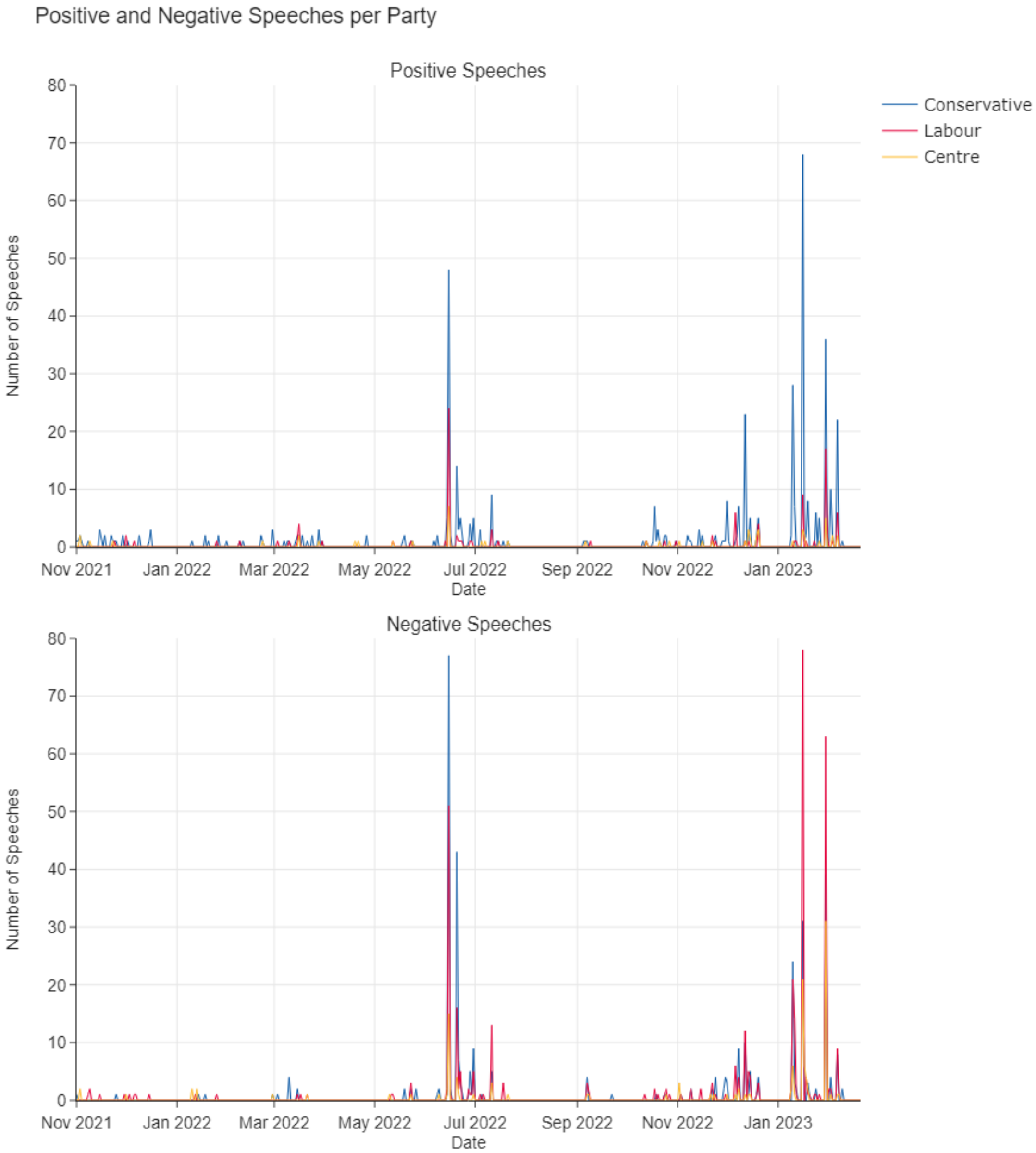

Figure 1: Number of positive and negative speeches per party

The centre parties, Liberal Democrat and the Scottish National Party, also have more negative than positive speeches about the strikes during the summer period. The apparent result that all three parties are more negative towards the strikes does align with how much disruption the strikes were causing and thus issues it was creating politically, as many people wanted something done about it.

The second period is during the end of 2022 and in early 2023, with the Christmas break visible. Again, many professions took industrial action during that period but also, the Conservative party introduced the Minimum Service Levels Bill, which was introduced towards the end of 2022. Starting in January, there is a very clear spike in negative speeches from Labour MPs, whereas the Conservative MPs start to speak

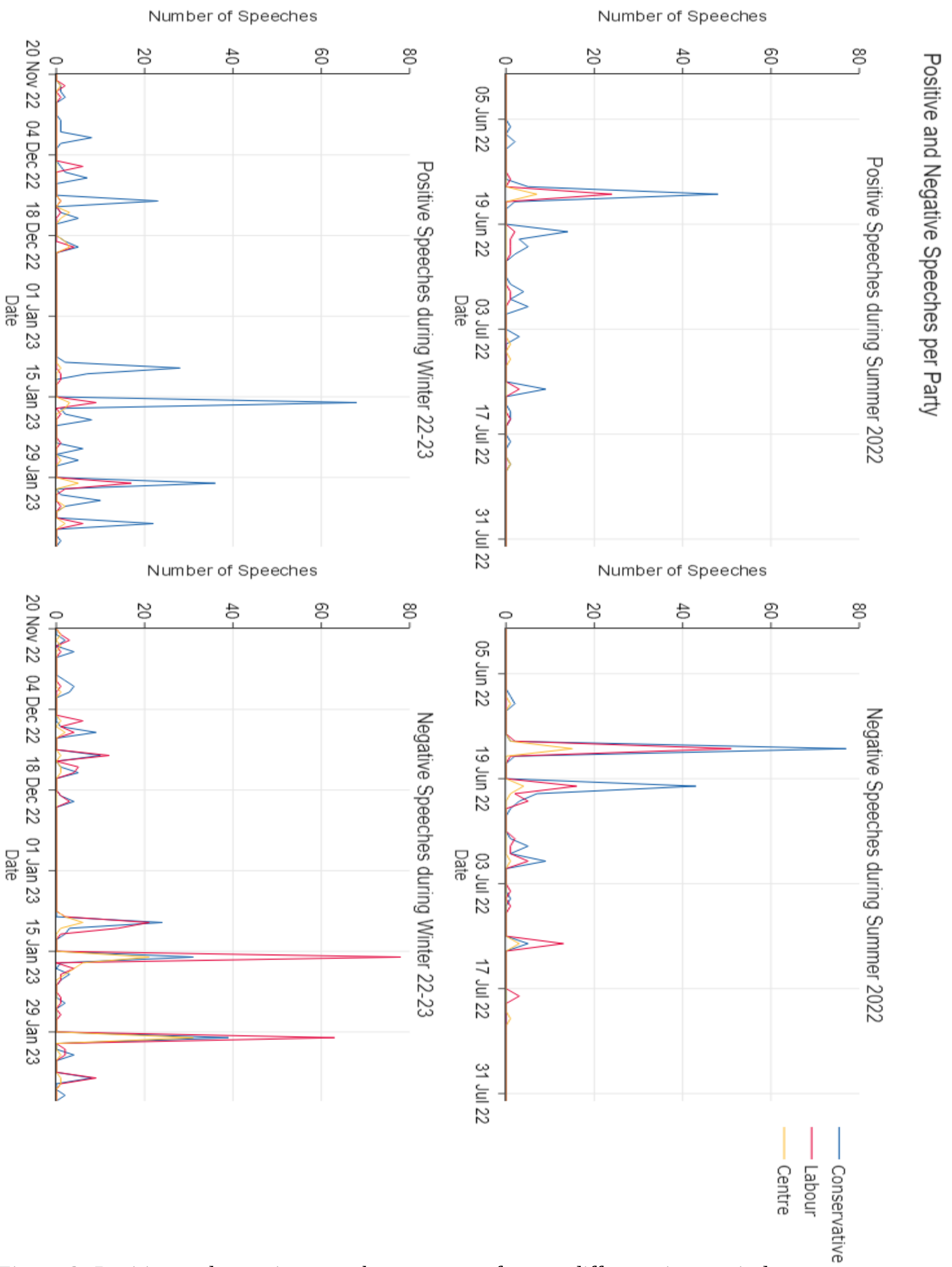

Figure 2: Positive and negative speeches per party for two different time periods

much more positively. Considering the introduction of the Conservative Bill, which the Labour Party strongly opposed, it could be that the mentions of strikes in the speeches in early 2023 are about the Minimum Service Levels Bill rather than the strikes themselves, explaining why the Conservative Party was more overwhelmingly positive compared to the summer of 2022 and why Labour MPs spoke so much more negatively.

In Figure 2 we can more clearly see the summer and winter periods of speeches. During the summer, there are two large spikes in both negative and positive speeches. These are days when there was a specific debate topic dedicated to the strikes, rather than when the strikes came up in debates about other topics. On 15th June, the House of Commons discussed the rail strikes and on 20th June, they discussed ‘Industrial Action on the Railway’. The increase in July likely was related to the debate on Employment Agencies and Trade Unions that took place on 11th July. In the winter, the House of Commons mostly debated the Strikes (Minimum Service Levels) Bill instead of the strikes themselves. A debate on the bill took place on the 16th and 30th of January, which are both very clearly visible in the graph (information about the dates of debates is available on the Hansard website hansard.parliament.uk/commons/).

6.1.2 Newspapers

To better compare the debates and the newspaper articles, the newspapers were aggregated by their political ideological leaning. The political ideology is based on previous research (see methodology chapter) and the categories include the following newspapers: right-wing (*Daily Mail* and *Mail on Sunday*, *The Daily Telegraph*, *The Sun*, *The Times* and *The Sunday Times*), left-wing (*Daily Mirror*, *The Guardian*), and centre (*The Independent*, *Metro*). Below in Table 4, the sentiment results are summarised for the three ideologies that will be used in further analysis.

Negative and Positive Articles per Newspaper Political Ideology

Political Ideology of Newspaper	Positive Articles	% Positive	Negative Articles	% Negative	Total Articles
Right-Wing	624	24.2%	1,954	75.8%	2,578
Left-Wing	431	32.2%	908	67.8%	1,339
Centre	1,038	40.1%	1,551	59.9%	2,589

n=6,506

Table 4: Negative and positive articles per newspaper political ideology

The right-wing news category includes more newspapers than the centre category, yet they published a roughly equal number of articles about the strikes. This indicates that the centre newspapers publish many more newspaper articles than the right-wing newspapers. Left-wing newspapers publish approximately half the number of articles that the other newspaper types publish. All three political ideologies published more negative than positive articles during the period from 1 November 2022 until 20 February 2023. 76 percent of all the articles published in right-wing newspapers were labelled as negative, with 68 percent for left-wing newspapers and 60 percent for centre newspapers.

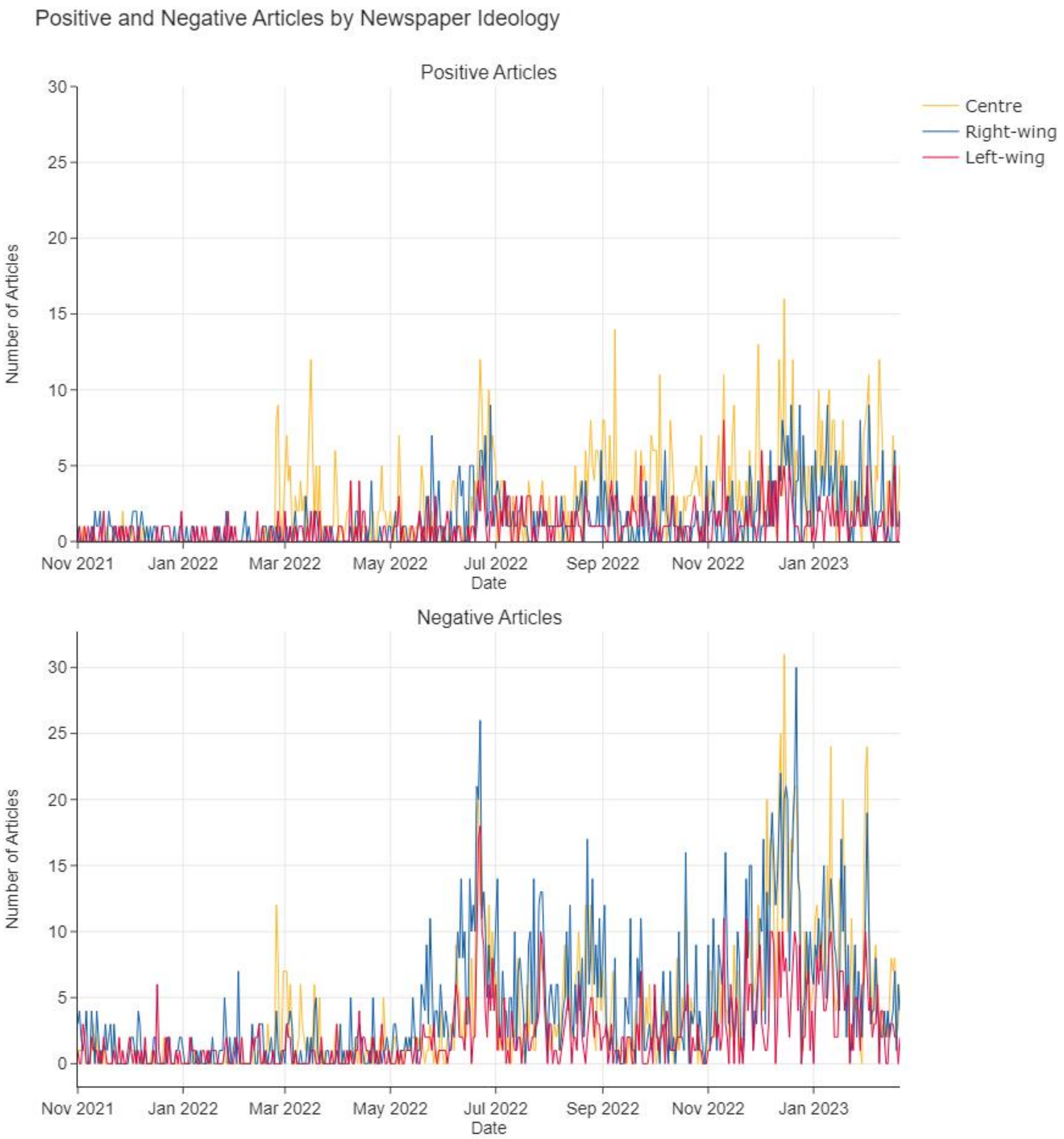

Figure 3: Positive and negative articles by newspaper ideology

In Figure 3, the sentiment in the newspaper articles is visible over time, split by newspaper ideology. The graphs clearly show that the centre newspapers publish more articles overall and have a relatively equal positive-negative split in their sentiment. This graph has more data points than the speeches graphs due to the frequent publication of articles. There is still a clear increase in articles published between late

2022 and early 2023. The spike in the speeches in the summer of 2022 is less clearly visible for the newspapers and there appears to be another period around March 2022 where more articles were published about the strikes. This period aligns with industrial action taking place across the UK, such as postal workers, barristers, education staff, and transport sector workers (Smythe, 2022), but did not appear to be much debated in parliament, with only a very small increase visible.

In the interest of comparison, Figure 4 shows the newspaper articles split up in the same time periods as the debate speeches above. In Figure 4, the increase in newspaper articles published during the specific periods of interest is more clearly visible. During the summer of 2022, there is a clear increase in the number of positive and negative articles on 15th and 20th June, when a debate on the strikes took place. In mid-December, there was a notable increase in articles seemingly unrelated to parliamentary activities. This could be attributed to the occurrence of a major rail strike and the ongoing industrial action by nurses (Austin, 2022; Triggle, 2022). There is also a clear increase in negative articles after the debate on the Minimum Service Levels bill was debated in the House of Commons on 30th January.

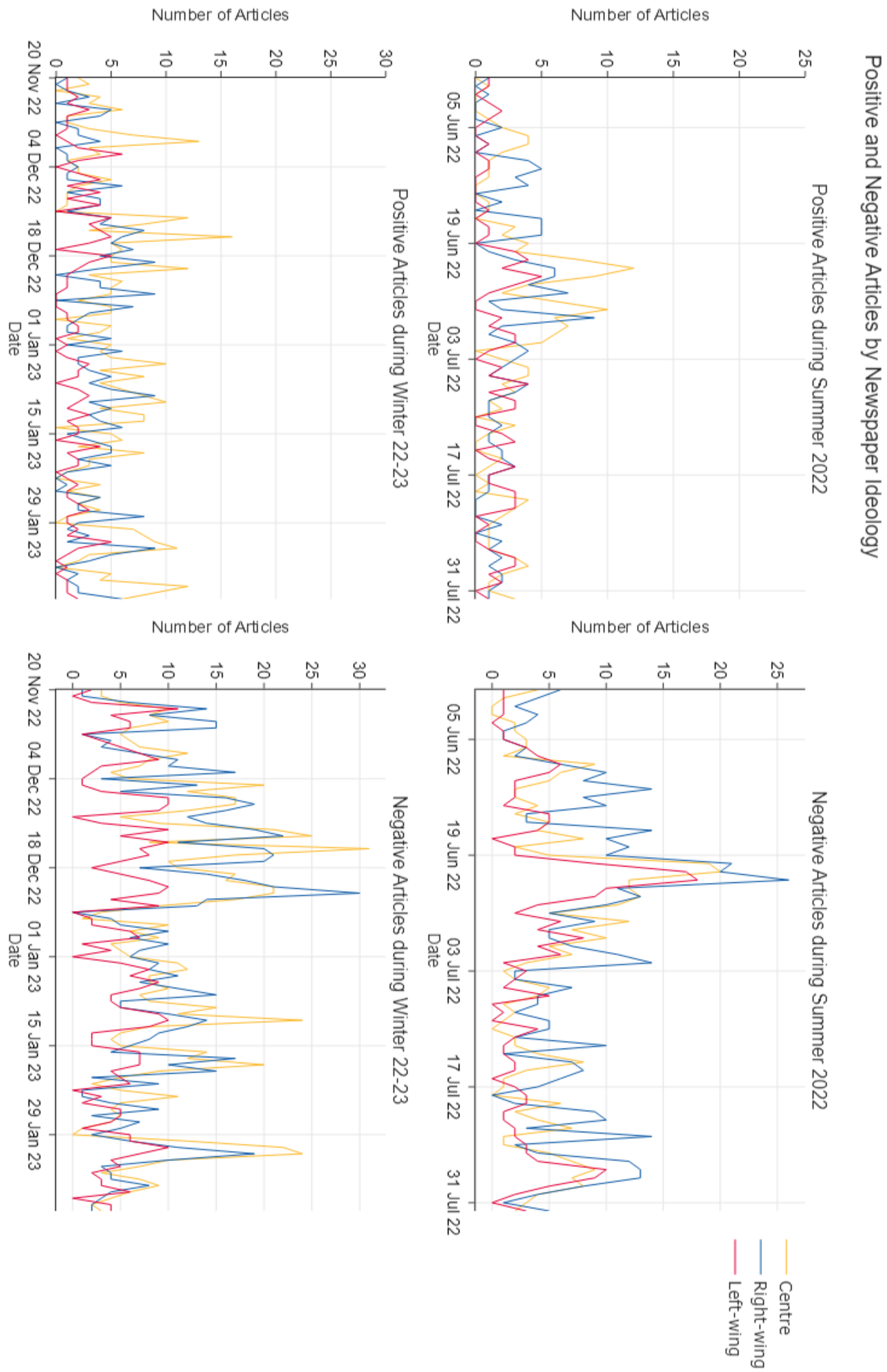

Figure 4: Positive and negative articles by newspaper ideology for two different time periods

6.1.3 Combined

In Figure 5 below, the speeches and articles can be compared to see the sentiment trends more easily. It appears that the debates and the news follow some of the same patterns when the number of speeches and articles goes up, especially when it comes to negative sentiment. It is apparent from the graphs that a spike in the debates is followed by a spike in the news. However, to investigate if the two influence each other, the next step of analysis is necessary: vector autoregression.

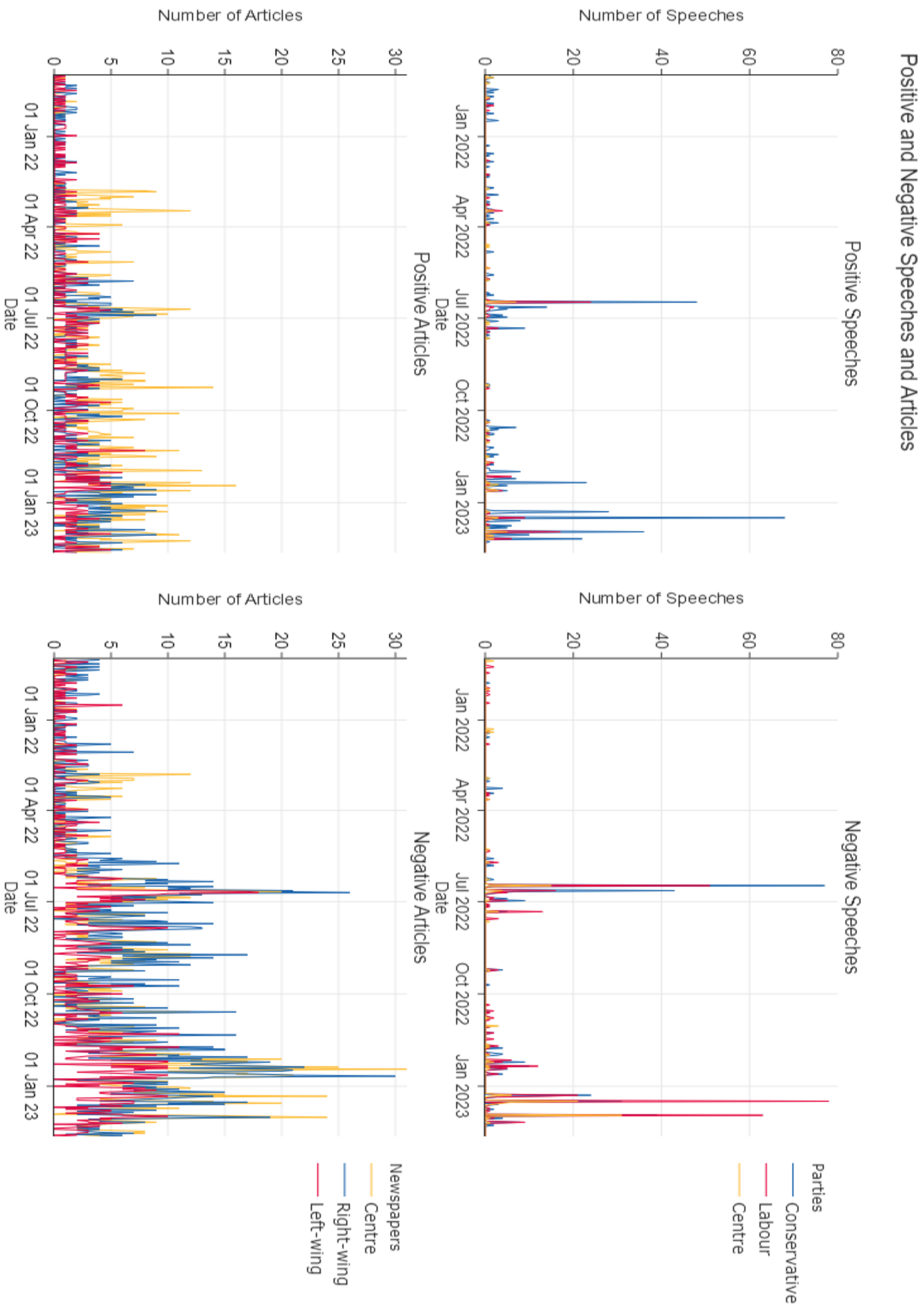

Figure 5: Positive and negative speeches and articles

6.2 VAR Results

In VAR analysis, each endogenous variable's present value is estimated using separate equations, considering both its past values and the past values of other variables (lagged values) (Vliegenthart & Montes, 2014, p. 327). Determining the suitable number of lags for each analysis is based on a comparison of models at a different lag and the corresponding fit statistics, such as the Akaike information criterion (AIC). Thus, the selection of lags in VAR is empirical and exploratory, within certain limits based on the model fit (Vliegenthart & Damstra, 2019, p. 26). In the case of this thesis, the effects of the variables might be delayed or manifest differently when including different additional variables such as political ideology. The estimation method for all the models was ordinary least squared (OLS). The output of a VAR model appears similar to the results of an OLS regression. Although the coefficients for each variable are provided in the model summary, they are challenging to interpret due to the possibility of multicollinearity caused by high correlations between lags of the same variable. Hence, alternative indicators are employed to assess the relationship between different time series: Granger causality tests and forecast error variance decomposition (FEVD) (Vliegenthart & Damstra, 2019).

6.2.1 Preliminary tests

Although the data is in the appropriate time series format, the data needs to meet certain assumptions for VAR models to fit correctly. For the tests and the VAR model itself, the 'statsmodels' library in Python was used (Seabold & Perktold, 2010). First, the time series was tested for seasonality, which both the speeches and the articles passed. Then, the time series need to be stationary, which means that the mean of the data remains constant over time (van Noije et al., 2008). The Augmented Dickey-Fuller (ADF) test was used to check for non-stationarity. The newspaper dataset was non-stationary so to render both datasets stationary, the difference between consecutive observations was taken. Taking the first difference changes how the time series looks and an example of the time series with the first difference taken is provided in Appendix A. Subsequent ADF tests showed that the transformed data was stationary.

Then, to indicate the suitable number of lags before the model is fitted, the partial autocorrelation of the variables was assessed. Partial autocorrelation plots indicate at which lag significant results might be found in the model. The graph for model 1 can be found in Appendix A. In the interest of parsimony, the model with the fewest significant number of lags is fitted (Vliegenthart & Montes, 2014, p. 330).

6.2.2 VAR Model 1

The first VAR model is fitted on the number of speeches and newspaper articles over time to create a baseline model to compare the more complex models to. To explore the potential relationship between speeches and articles, the Pearson correlation coefficient (r) is considered. The coefficient $r = 0.21$ indicates a moderate relationship between the two variables. However, this does not mean that a causal relationship is present. The results of the VAR analyses, which examine the lagged values of the variables, will estimate this.

The partial autocorrelation plots for both variables suggest a lag of six. After fitting the model, the AIC was used to determine that the model with six lags was indeed the best one because it had the lowest AIC score. AIC was chosen for this thesis because this score penalises complex models, which thus prevents overfitting in the later models. The summary (see Table 5) of this model suggests that a lag of three is the first significant option for articles to impact speeches and a lag of one for the speeches to impact the articles. Using the significant lag from the model, the Granger causality tests demonstrate that this effect is present for the debates and for the news, meaning that articles, lagged three days, influence the number of debate speeches and that speeches, lagged one day, influence the number of news articles.

VAR Model 1

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Speeches						
Constant	0.03	0.88	0.03	0.98 n.s.		
Articles $_{t-3}$	-0.24	0.12	-2.04	0.04 *	0.00 ***	2.4%
Articles						
Constant	0.08	0.38	0.21	0.83 n.s.		
Speeches $_{t-1}$	0.08	0.02	4.12	0.00 ***	0.00 **	3.2%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 5: VAR Model 1 summary

The FEVD shows that changes in the news (lag 3) account for 2.4 percent of the variation in parliamentary speeches and shocks in the parliamentary speeches (lag 1) account for 3.2 percent of the variation of the news.

The model was tested for normality using the 'statsmodels' Python package. The model violated this test, which is likely due to the low means in the parliamentary debate data. However, VAR models, because they are based on OLS estimations, are robust against violations of normality (Vliegenthart & Montes, 2014).

6.2.3 VAR Model 2

The next VAR model has sentiment added to determine if the tone of the speeches and articles plays a role in the relationship between the two. The correlation between the variables is similar to the first model, with $r = 0.20$ between the negative news and the negative speeches and $r = 0.19$ for the positive variables. Positive news correlates more weakly with negative speeches with $r = 0.15$, but negative news correlates more with positive speeches with $r = 0.20$. The partial autocorrelation plots indicated six lags and the AIC for the model with six lags was the smallest.

VAR Model 2

Articles	*Negative speeches*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.02	0.57	0.03	0.97 n.s.		
Negative $_{t-6}$	-0.20	0.09	-2.16	0.03 *	0.03*	2.4%
	Positive Speeches					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.00	0.32	0.01	0.99 n.s.		
Negative $_{t-6}$	-0.10	0.05	-1.93	0.05 n.s.	0.01 *	2.9%
Speeches	*Negative Articles*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.05	0.29	0.16	0.88 n.s.		
Negative $_{t-1}$	0.36	0.11	3.35	0.00 **	0.00 ***	4.5%
	Positive Articles					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.03	0.16	0.18	0.85 n.s.		
Negative $_{t-1}$	0.13	0.04	2.96	0.00 **	$_{(t-2)}$0.04 *	0.5%
Positive $_{t-1}$	-0.20	0.08	-2.55	0.01 *	0.83 n.s.	0.8%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 6: VAR Model 2 summary

Table 6 contains the model summary, with all significant predictor variables added for each outcome variable. The lag suggested by the partial autocorrelation plots aligned with the significant result in the model, except for positive speeches. This variable had no significant predictor, so negative articles (lag 6) was added because it had the lowest p-value. Interestingly, the Granger causality test reflected the relatively high correlation and indicates that negative articles Granger causes positive speeches.

In this model, we can see that the significant lag of the model does not always line up with significant lags in the results of the Granger causality tests. This difference is because although VAR models are fitted on all variables, Granger causality tests are only conducted on two variables at a time. Therefore, the significant variables indicated by the model summary are used to determine which lag to use for the Granger causality and the FEVD. For example, in the case of positive articles, the Granger causality test for negative speeches was not significant at lag one but it was significant at lag two. In the interest of understanding the relationship between the variables even at a different lag than was significant in the model, a significant lag from the Granger causality test was added to the table if present.

The FEVD scores are similar to the first model, which means the size of the effect of the variables on each other did not change. For negative speeches, negative news (lag 6) accounted for 2.4 percent of the variance, and for positive speeches negative news (lag 6) accounted for 2.9 percent of the variance. For the news, negative speeches (lag1) accounted for 4.5 percent of the variance in negative articles and negative and positive speeches (lag 1) accounted for 0.4 and 0.8 percent of the variance in positive articles, respectively. For the FEVD plot, see Appendix B.

6.2.4 VAR Model 3

Now, to understand the sentiment in the speeches and the news as well as more complex relationships between the news and debates, the political party of the MP speaking and the political alignment of the newspapers were added to the third VAR model. A heatmap with all the correlation coefficients can be found in Appendix B. Between the news and speech variables, the strongest correlation can be found between positive Conservative speeches and negative left-wing and centre news articles (r = 0.26). Between negative Conservative speeches and negative left-wing and centre news articles, the correlation is similar (r = 0.24).

Based on the partial autocorrelation plots, a maximum of seven lags was chosen. The AIC confirmed that seven lags was the best fit. Due to the large number of variables in this model, the results below only show the relationships between any news and any debate variable where a significant p-value could be found. As the focus of this thesis is on the relationship between news and parliament, any significant inter-party or inter-newspaper relationship was omitted. Additionally, for each significant variable, the lowest significant lag was chosen for that variable. Due to the number of variables in this model, Table 7 and Table 8 below only show the relationships with the highest FEVD score. For the full table of results for VAR model 3, see Appendix C.

VAR Model 3 Part 1 (Speeches)

Articles	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
	Negative speeches – Centre parties					
Negative Centre $_{t-1}$	-0.08	0.04	-2.08	0.04 *	$_{(t-2)}$0.00 ***	2.8%
	Positive Speeches – Centre parties					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Right-wing $_{t-3}$	0.03	0.01	2.18	0.03 *	0.00 ***	1.1%
	Negative speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre $_{t-3}$	-0.33	0.13	-2.55	0.01 *	0.00 ***	1.3%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre $_{t-2}$	-0.22	0.11	-2.08	0.04 *	0.00 ***	3.7%
Negative Left-wing $_{t-4}$	-0.36	0.17	-2.09	0.04 *	$_{(t-3)}$0.00 **	3.4%
	Negative speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre $_{t-2}$	-0.35	0.13	-2.78	0.01 **	0.00 ***	3.3%
	Positive speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-5}$	0.16	0.08	2.08	0.04 *	0.00 ***	1.2%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 7: VAR Model 3 for speeches

Overall, the significant variables in the model are primarily also Granger causal. The following newspaper article variables most often have a significant influence on the speeches: negative centre, negative left-wing, negative right-wing, positive right-wing, and negative left-wing. Negative left-wing news is often significant at lag one, both for negative speeches by Labour and Conservative MPs. The positive and negative speeches by the Centre parties and the negative Labour speeches are influenced by news at a lower lag. On the other hand, the news that affects Conservative speeches is significant at higher lags.

The FEVD for centre newspaper articles on speeches is the highest on the positive speeches by Conservative MPs (3.7%) and on negative Labour speeches (3.3%). Negative left-wing articles have a similar effect on negative Conservative speeches (3.4%). The negative centre newspapers account for the most variance in the speeches of all the parties.

VAR Model 3 Part 2 (Articles)

Speeches	*Negative articles – Centre newspapers*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre parties $_{t-1}$	0.48	0.21	2.23	0.03 *	0.04 *	2.8%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre parties $_{t-7}$	-0.53	0.19	-2.79	0.00 **	$_{(t-3)}$0.03 *	2.6%
	Negative articles – Right-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Conservative $_{t-5}$	0.25	0.11	2.21	0.03 *	$_{(t-3)}$0.00 **	2.8%
	Positive articles – Right-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Centre $_{t-4}$	-0.35	0.17	-2.02	0.04 *	0.02 *	3.7%
	Negative articles – Left-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Conservative $_{t-3}$	0.15	0.07	2.36	0.02 *	0.00 ***	1.4%
	Positive articles – Left-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Conservative $_{t-2}$	-0.12	0.05	-2.34	0.02 *	0.73 n.s.	1.3%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 8: VAR Model 3 for articles

Table 8 contains the results for the newspaper articles (full table in Appendix C). Similar to the results for the speeches, the articles each have four significant predictor variables. The speech variables that were most often significant are: negative Centre, negative Conservative, positive Labour, and positive Conservative. These variables are significant for five outcome variables. Except for negative left-wing news, the Granger causality tests line up with the significant variables in the model. For all newspapers except for the left-leaning ones, it is the Centre speeches that account for the most variance. For the centre newspapers, the FEVD score is 2.8 percent for negative and 2.6 percent for positive articles. For the positive right-wing newspapers, the FEVD value is 3.7 percent.

This model is very complex and seeing the results in a table does not provide a clear overview of which relationships are present in terms of the number of positive and negative speeches and articles affecting each other. The network graph (Figure 6) below shows which variables affect each other. An arrow was drawn if, at the same lag, the variable was significant in the VAR model and also had a significant Granger causality test. Many of the relationships between the speeches and the news are bidirectional. The speeches are also influenced by more than one newspaper and sentiment. The node for positive left-wing news is missing entirely, which means that this node had no relationships that were significant in the VAR model and the Granger causality tests. The positive centre news had no outward arrows, meaning that it does not significantly influence any speeches. The same applies to the positive and negative Centre speeches.

Network of The Relationships Between Newspaper Articles and Debate

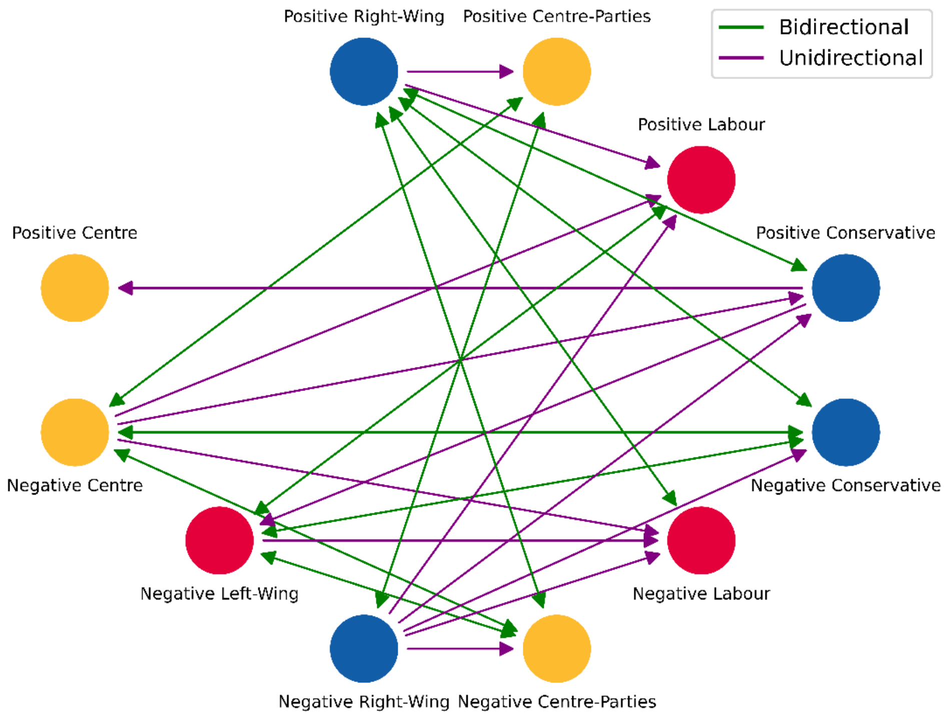

Figure 6: Visualisation of VAR model 3.

6.2.5 VAR Model 4 and 5

In the time series graphs (Figure 5), there are two distinct periods visible with increased news and increased debate about the strikes. Not only are there significantly more articles and speeches, but there also appears to be a shift in the tone of the speeches of the Labour and Conservative parties. Therefore, two more VAR models were fitted to determine if there is a difference in the relationship between articles and speeches. During the summer, it is probable that the news and debates were focused on the strikes, whereas during the winter, they might have centred around the Conservative strikes Bill. The full tables can be found in Appendix D and E, the tables below only contain the significant variables with, if possible, also significant Granger p-values, and the highest FEVD.

VAR Model 4 (Summer 2022)

Speeches	*Negative articles – Right-wing Newspapers*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Labour $_{t-2}$	-7.54	2.60	-2.90	0.00**	$_{(t-1)}$0.05 *	6.1%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Conservative $_{t-2}$	-0.91	0.42	-2.16	0.03 *	0.70 n.s.	4.9%
Articles	*Negative speeches – Centre parties*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-1}$	0.42	0.18	2.37	0.02 *	0.17 n.s.	9.0%
	Positive speeches – Centre parties					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-1}$	0.17	0.08	2.10	0.04 *	0.26 n.s.	10.3%
	Negative speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-1}$	1.73	0.64	2.69	0.01 **	0.24 n.s.	8.7%
	Positive speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-1}$	0.67	0.27	2.51	0.01 *	0.35 n.s.	8.3%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-1}$	1.48	0.59	2.49	0.01 *	0.16 n.s.	9.1%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 9: VAR Model 4 summary

For model 4, the period is 30 May until 1 August 2022. As visible in Table 9, not all variables have significant results. Also, the negative left-wing newspaper articles seem to have the most impact on the debates for all three parties. There are fewer significant causal relationships and the FEVD values are much higher than in the large model.

VAR Model 5 (Winter 2022-3)

Speeches	*Negative articles – Right-wing Newspapers*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Labour $_{t-2}$	-1.50	0.66	-2.26	0.02 *	0.10 n.s.	4.4%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Labour $_{t-1}$	-1.82	0.89	-2.04	0.04 *	$_{(t-3)}$0.04 *	3.8%
	Positive articles – Left-wing newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Labour $_{t-1}$	-0.58	0.28	-2.05	0.04 *	0.99 n.s.	2.9%
Articles	*Negative speeches – Centre parties*					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-2}$	1.02	0.38	2.65	0.01 **	0.44 n.s.	37.1%
	Positive speeches – Centre parties					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-2}$	0.19	0.07	2.67	0.01 **	0.08 n.s.	28.7%
	Negative speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-1}$	-1.79	0.83	-2.16	0.03 *	0.50 n.s.	17.6%
	Positive speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-1}$	-0.53	0.22	-2.48	0.01 *	0.56 n.s.	19.1%
	Negative speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Positive Right-wing $_{t-2}$	-1.60	0.75	-2.13	0.03 *	0.71 n.s.	20.3%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Negative Left-wing $_{t-2}$	-2.77	1.24	-2.23	0.03 *	0.17 n.s.	16.4%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Table 10: VAR Model 5 summary

For model 5, the period is 29 November 2022 until 10 February 2023. In Table 10, the news has more of an influence on the debates than vice versa when looking at the number of significant variables in the model. The number of positive and negative Labour speeches influences the number of newspaper articles in all three ideological leanings. Few newspaper articles had a significant causal effect on any of the speeches based on the Granger causality tests. The FEVD values for positive right-wing newspaper articles are very high in this model, with the FEVD of the articles on negative Centre Party speeches reaching 37 percent. This means that the variance in these articles was accounted for to a great extent by the negative speeches by Centre MPs.

6.3 Analysis

A summary table with the most important finding of each model can be found in Appendix F. Table 11 contains an overview of all the hypotheses and whether they have been accepted based on the results.

Overview of Hypotheses

No.	Content	Accepted?
1	Newspaper articles and parliamentary debate speeches have an effect on each other's frequency.	✓
1a	The parliamentary debates have a delayed reaction to the news, whereas the news responds quickly to the parliamentary debates.	✓
2	The type of sentiment expressed in a speech or news article influences the relationship between the news and parliament.	✓
2a	The number of negative newspaper articles affects the number of negative and positive parliamentary speeches more than positive articles.	✓
2b	The number of negative speeches affects the number of negative and positive newspaper articles more than positive speeches.	✓
3a	Conservative MPs will express more negative sentiment in the speeches about strikes than their Labour and Centre peers.	X
3b	The influence of distinct newspaper ideologies and their sentiment varies across the speeches of different political parties.	✓
4a	Right-wing newspapers will express more negative sentiment in the articles about strikes than their centre and left-wing counterparts.	✓
4b	The influence of political party alignment of the speeches and their sentiment varies across the articles of different newspaper political ideologies.	✓

Table 11: Overview of hypotheses

6.3.1 VAR Model 1

In VAR model 1, the Granger causality tests were statistically significant in both directions, which indicates that the news and the parliament indeed affect each other. Based on the FEVD, the effect is comparable with 2 per cent for articles on speeches and 3 percent for speeches on articles. So, although the number of articles affects the number of speeches and vice versa, the explained variance is minimal and these variables are still mostly explained by their own past values. Additionally, no conclusions can be drawn about which one affects the other more because the percentages are very close.

The above results are in line with existing research about agenda setting in the UK, which states that, as opposed to other countries, there is still a significant influence of politics on the news (Vliegenthart et al., 2016). However, Vliegenthart et al.'s research did still find that the media's effect on parliament was greater than the other way around. Their research used parliamentary questions instead of parliamentary debates, which are more flexible and thus adaptable to the news, whereas debate agendas must be set a few days before. Regarding the news, they only selected one very influential newspaper rather than a range of different newspapers representing varying political leanings.

Furthermore, the difference in findings could be due to the topics that they included. According to issue typology theory, the strikes are not strictly a sensational issue which would attract more media attention (Soroka, 2002). Therefore, because the strikes are less phenomenal than issues such as international security and more of an economic and social policy issue, the parliament had more influence on the news than on other issues. Based on this theory, the results of this thesis are in line with what was expected for this specific issue.

When looking at the lags in model 1, there is a difference between when the newspaper articles significantly affect the debate speeches and vice versa. The number of articles about the strikes can be partially accounted for by the number of speeches that mention strikes the day before. The different lag values suggest that the news responds to parliamentary debates more quickly, after one day, than the parliament does to the news, after three days. Considering that an important role of the news is reporting on the happenings of politics in the country, this is an important finding. Also, keeping in mind that the House of Commons determines debate agenda points ahead of time and does not meet on weekends, it makes sense that the parliament would not talk more about the strikes after it has been in the news until a few days after the fact.

The results related to the difference in how long it takes for the news to influence politics and vice versa are in line with van Noije et al.'s (2008) theory that the parliament has a delayed reaction to the media because of its more rigid agenda that is set a few days before a debate takes place. The news on the other hand is supposed to report on the happenings of the day before.

6.3.2 VAR Model 2

In VAR model 2, as outlined above, the main relationship stays significant, with speeches significantly influencing articles and vice versa. The Granger causality tests were also significant both ways. Thus, the results of model 2 also support the expectation that the effect between news and parliament is bidirectional.

Compared to the first model, the parliament speeches appear to have an even slower reaction, going from 3 to 6 lags for both negative and positive articles. This difference in lags does not counter the theory, however, as there was no expectation for specific days but rather that the parliament would take longer to be influenced by the news than the other way around. The lag results of this model for the news are the same as in VAR model 1, which reinforces the expectation that the news responds very quickly to what is being said in the House of Commons.

However, with the added layer of sentiment, only negative articles are significant in predicting the negative and the positive speeches according to the model p-values and the Granger causality tests. For the news, it is again the negative speeches that are significant in predicting both positive and negative articles. The difference in the FEVD between models 1 and 2 also indicates that the sentiment does play a role in the relationship between the news and parliament. The FEVD for negative news on negative and positive speeches is similar to that of the first model. Adding sentiment did cause a bigger change for the news variables. In model 2, the FEVD for negative speeches accounting for variance in the negative articles was a decent bit larger than in model 1. However, the effect of negative and positive speeches on the positive news articles was reduced to below 1 percent. Thus, the FEVD scores suggest that the negative speeches account for the most variance of any variable and that effect is on negative articles.

The finding that negative news has more influence than positive news is in line with the theory that politicians are more sensitive to critical and negative news (Vliegenthart & Montes, 2014). As expected based on the previous research about negativity in the House of Commons, the negative elements of the debates were more influential (Finlayson, 2017). In this thesis, the negative speeches had more of an effect on the news than positive speeches.

6.3.3 VAR Model 3

6.3.3.1 Descriptive statistics

The descriptive statistics for the Hansard data show that the Conservative party is the only party that has a majority of positive speeches. At first consideration, this seems unusual because traditionally the Labour Party is more pro-union and pro-strike (Davies & Nophakhun, 2018). However, keeping in mind their Minimum Service Levels bill proposal in early 2023, they could speak more positively about the policy proposal related to the strikes. The high number of positive speeches by Conservative MPs, especially in early 2023 when there was not also a large number of negative speeches like in the summer of 2022, was likely in reference to the Strikes Bill. So, instead of interpreting the result as contrary to theory, it can also be considered as the Conservative Party being positive towards a policy that they themselves introduced. In turn, the Labour Party was negative towards a Bill that was very anti-strike, which is in line with their party identity. Additionally, Milner's initial assessment of the recent strikes found that the "Labour response has been muted" and that "The party's unhappy divisions over how far to support workers' rights, pay and conditions were in full evidence in late July" (2022, p. 43).

Because the differences are so clear between the different parties, the results of the sentiment analysis suggest that there is indeed a relationship between the party of the MP and the sentiment expressed in the speech.

The descriptive statistics for the newspaper data indicate that all three ideologies are negative in their coverage of the strikes. This aligns with previous research suggesting that newspapers are negative towards strikes (Gillespie, 2021). What is more, the right-wing newspapers are a lot more negative than the others, which also reaffirms previous research that right-wing newspapers are anti-strike (Macaulay, 2016). The left-leaning newspapers also published a majority of negative articles, which could potentially also be impacted by their coverage of the Conservative strikes Bill. The centre newspapers were the most balanced when it came to the sentiment expressed in the articles, which aligns with their desire to publish more neutral and balanced news.

In the news, the two time periods outlined in the discussion of the speeches are also present. However, the peak for the newspaper articles in winter 2022-3 is slightly before the dates that the strikes Bill was discussed in parliament. Therefore, it could be that the newspaper sentiment during the winter of 2022-3 is towards the strikes, not the strikes Bill. This would explain why there is no clear shift in the sentiment being expressed by the different newspaper ideologies besides an increase in the number of articles being published. Yet, looking at the zoomed-in graph in Figure 4, there are some spikes in negative coverage around the 16th and the 30th of January 2023, when the bill was being discussed, with no spike in positive coverage to balance it.

Based on the variation in the proportion of negative and positive articles published, the sentiment analysis results suggest that the political leaning of a newspaper impacts the tone expressed in the articles.

6.3.3.2 VAR Results

The results of the models with more variables are meant to unravel the complexities of the relationship and add another layer of understanding. In VAR Model 3, the party of the MP and the political ideology of the newspaper were added. Considering Figure 6, there are many cross relationships, meaning that, for example, positive speeches are affected by both negative articles and different parties are affected by different newspaper ideologies. In the graph, it is also again very clear that negative news influences speeches more than positive news.

When considering the more elaborate results in Table 7, the Centre parties and the centre newspapers had fewer significant variables affecting them, which indicates that these MPs and newspapers are less swayed by either the news or parliament. Overall, the results suggest that it is not a straightforward one-to-one relationship, and that negative news from certain politically aligned news outlets is more likely to influence MPs than other news articles. MPs seem to respond the quickest to negative left-wing news with the lag often being significant at 1, both for negative speeches by Labour and Conservative MPs.

The FEVD values suggest that the negative speeches by the Centre party, positive Conservative speeches, and negative Labour speeches have the most variance accounted for by newspaper articles. This means that these speeches are most influenced by what is being said in the news. Additionally, centre newspaper articles account for the most variance, which means that the sentiment expressed in these articles most often influenced the sentiment expressed in speeches. Notably, the negative speeches of all parties tend to have higher FEVD values than the positive speeches, which indicates that MPs are more influenced by the news when they are speaking about the strikes negatively. The Conservative party also had high FEVD scores for positive speeches, which suggests that their speeches are generally quite influenced by the news. It is of note that in this model, the lags are much more variable than in the previous model. This suggests that by adding the political ideology of the newspapers, the time it takes for the MPs to be affected by the news varies more . It indicates that negative news affects MPs faster than positive news and that articles from right-wing newspapers often take longer to influence MPs, reinforcing hypothesis 2 that negative news is more influential.

As visible in Figure 6, the negative right-wing newspaper articles are not significantly affected by any speeches based on the VAR model p-value and Granger causality tests. Even the negative Conservative speeches do not influence the negative right-wing articles on both significance criteria. This is notable and could be because the negativity expressed is related directly to the strikes rather than what is being debated in the House of Commons. This indicates that the right-wing newspapers might not be as politically informed as the other newspapers when it comes to writing their negative articles. The positive left-wing newspaper articles node is not present at all. This could mean that these articles are less influenced by parliament than negative left-wing articles and articles from other newspapers.

By adding political alignment, the size of the effect of parliament on news is reduced. Only a few relationships still maintain high FEVD values. Yet, keeping in mind Granger causality, the parliament does affect all the newspapers, except for the positive left-wing articles. In terms of sentiment, it is more balanced whether positive or negative speeches affect the articles. Both negative and positive speeches seem to have a similar effect on negative and positive newspaper articles, which is not strictly tied to which party the MP belongs to. For example, for the positive right-wing articles, the negative Conservative speeches have a significant effect, and for the negative left-wing articles, the positive Labour speeches have an effect. This reflects the diversity of how the strikes are discussed in parliament by the different MPs, although they belong to certain parties, they do have individual opinions.

The findings indicate that in some cases, the sentiment expressed in a speech might not line up with the sentiment it is affecting in the newspaper article. For example, the negative Conservative speeches affecting negative left-wing articles is not a reflection of similar sentiment but most likely the left-wing newspapers criticising that the

Conservative party was negative about the strikes, rather than echoing the negativity towards the strikes. For left-wing newspapers, the insignificant result of all the Granger causality tests indicates that these articles are not easily influenced by what goes on in the parliament when it comes to their positive articles. This means that these articles are most likely directly positive about the strikes rather than talking about what was discussed in parliament related to the strikes.

The FEVD scores in this model are lower than for newspaper articles in the previous model, which could be because the explained variance was mostly happening in certain types of newspaper articles. Based on this, the strongest effect of the speeches in parliament is on the centre newspapers.

Overall, it is notable that both negative and positive speeches by all three parties had more than one significant relationship to the news, which again indicates that there is a relationship between the news and parliamentary speeches.

6.3.4 VAR Model 4 and 5

VAR models 4 and 5 explore the differences in how the parliament and news affect each other when there is a clear policy issue at hand. In the summer period, when many strikes were happening at once, the parliament was more influenced by the news than vice versa. During this period, only the negative right-wing newspapers were affected by speeches (positive Labour), whereas all the speeches were affected by at least one newspaper variable, most often negative left-wing articles. The MPs could thus be using the newspapers, the most prominent information source, to inform their speeches about the strikes. During the winter period, when the Conservative party formally introduced their strikes Bill, the news appeared to have less of a causal influence on the parliament. This could be because the MPs were speaking more about the policy issue at hand with the Bill rather than the strikes themselves.

6.4 Limitations

6.4.1 Data

When it comes to the data, choosing to use parliamentary debates as a representation of politics is a less flexible choice than using parliamentary questions as used in other research into the relationship between media and politics (Vliegenthart & Montes, 2014). In the case of debates, the agenda must be set in advance, whereas parliamentary questions are more *ad hoc*. However, by using parliamentary debates, what is being said has the potential to influence policy, whereas using questions might increase the likelihood that certain newsworthy topics come up but there is no direct impact of it being discussed by politicians. Furthermore, newspaper articles are an imperfect representation of the mass media. The specific publications selected for this thesis might not be fully representative of the UK population, and with the advent of social media, many people are getting their news from other sources.

Additionally, the way the data was filtered could have introduced errors. The method of selecting which speeches to include in the dataset is not without its flaws because there is no differentiation between different strikes, a mention of a strike or a law about strikes. The same goes for the newspaper articles, although the filtering method was as rigorous as possible, there could still be newspaper articles in the dataset that do not actually discuss the strikes in the UK. However, this limitation is moderated by the large sample size.

The chosen method of aggregating related to political party and ideology is also not without its drawbacks. Sorting the Scottish National Party and the Liberal Democrats into a single 'Centre parties' category does not fully represent the political nature of these parties. Political alignment is a spectrum, and categorical variables cannot entirely accurately reflect this. In addition, the categories chosen for the newspapers are not uncontested, as newspapers are not fully uniform in their political leaning.

When it comes to the number of speeches per party and the number of articles by newspaper type, there is an imbalance there. However, this imbalance is most likely representative of the real world, where the Conservative party does speak more in the House of Commons because they have more MPs and the centre newspapers included in this thesis do actually just publish more articles about the strikes. Because the VAR model is fitted on frequency counts over time, the chosen methodology takes care of this imbalance.

6.4.2 Method

6.4.2.1 Sentiment analysis

Despite recent strides in the accuracy of NLP, especially deep-learning models, they are still only an approximation of human readers. The chosen model, SiEBERT from Hugging Face, could have mislabelled speeches and articles. However, this limitation is addressed by having larger sample sizes. It is also overshadowed by the benefits such as reducing human bias and the ability to create a reproducible analytical pipeline (Indulska et al., 2012). At a lower level, the circumvention of the 512 token limit introduced by Hugging Face's truncation could have imperfect label outputs because longer texts are not analysed holistically. However, even state-of-the-art language models such as GPT-4 cannot get around these limitations (Yang et al., 2023).

6.4.2.2 VAR

VAR models are incredibly difficult to interpret, especially with many variables and a large maximum lag. VAR, like any other statistical method, can only say so much about the actual relationship between the variables. In terms of inherent limits, VAR models do not allow for instantaneous causation. Further research could use structural equation models with seasonal influence, short-term influence and instantaneous causation.

Furthermore, as with any statistical model, there may be unobservable factors that may confound the results that are not included in the VAR model (Barberá et al., 2019). Additionally, the Granger tests do not indicate the direction of the causal relationship. For example, it is unclear whether the increase in one variable Granger causes an increase or a decrease in another variable.

The focus was on the variables between newspapers and speeches, not inter-relationships. So, for example, the Conservative speeches could have had a significant effect on the Centre Party's speeches. But, keeping in mind the central question of this thesis, only the relationship between the news and parliament was considered.

6.4.3 Topic

Choosing to use strikes as the case study for investigating the relationship makes interpreting the results more complex. Positive and negative sentiments can mean different things when talking about strikes because it is a multifaceted issue, which is then further complicated by political ideology. For example, when the Labour Party speaks negatively, it could be that they are being negative in response to something that a Conservative MP said rather than being negative about the strikes themselves. What is more, when a positive article is published about the strikes, it could be in reference to being relief that it is over or on the other hand it could be expressing support for the success of the strike. However, in current UK affairs, the issue of strikes covers much ground. It is related to inflation and the cost-of-living crisis, some of the biggest problems facing the UK population at the moment. It is also an issue which is regularly represented in both the news and in parliament, which was an important requirement for answering the central question in this thesis. This could potentially be addressed by using better entity-focused sentiment analysis to get a measure of sentiment expressed towards a specific entity, rather than the sentiment of a text as a whole. Nevertheless, even cutting-edge entity sentiment analysis cannot understand abstract concepts such as strikes, strikes of a specific profession or strikes happening during a specific time period.

Generally speaking, the findings of this thesis are only applicable to the case of the UK. Previous research has indicated that the effect of the news on parliament is different depending on the country-specific context, even within Europe. The thesis has provided the means for this research to be replicated in other countries with all the necessary code available on GitHub.

7 Conclusion

The results of this thesis are in line with previous research that the relationship between the news and parliament in the UK is bidirectional. In other countries, the parliament was not found to have a strong influence on the news, making the UK stand out (Vliegenthart et al., 2016). In the context of the recent strikes in the UK, a hotly debated issue in the news and parliament, the results of this thesis empirically demonstrate that the number of newspaper articles published about the strikes affects the number of speeches in parliament about the strikes, and vice versa. The findings also reveal that there is a relationship between the sentiment expressed in parliament and in the news, which is influenced by political alignment.

Based on the results, the articles and the speeches account for a similar amount of variance in the other, which means that neither has more power. Therefore, contrary to previous research, in the case of strikes, the news does not dominate (van Noije et al., 2008). Additionally, the findings indicate that the effect between the variables is independently lagged. News articles take longer to influence parliamentary speeches than the other way around. This finding is critical as it shows how incredibly important the news media are for the public in how they portray debates on issues of public importance. Because the news reports so quickly on parliament, it is likely that the public's perception is shaped by the newspaper framing rather than how the politicians discuss the matter. This was to be expected, as not many people watch the debates in the House of Commons or read the transcripts. Additionally, the findings indicate that the sentiment expressed by MPs is not always reflected in the news, which reinforces the idea that the news frames what is said in parliament. However, when there is a clear policy issue being discussed in parliament, the MPs tend to be less influenced by the news.

This thesis has also investigated additional factors that play a role in this relationship between the news and parliament by adding sentiment and political alignment. The findings indicate that the association between news coverage and the speeches of MPs is indeed incredibly complex. It appears that negative news from specific news outlets aligned with certain political views is more likely to impact MPs than other news. For the newspapers, it is also the negative speeches that are more likely to influence the articles. Therefore, sentiment analysis is crucial when studying the relationship between news and parliament because it helps us understand the underlying emotions and opinions expressed in news articles and parliamentary speeches. By analysing the sentiment of these texts, we can gain insights into the stance of politicians and newspapers on specific issues and how they may be influencing public opinion.

The descriptive statistics of the sentiment data also reveal interesting information about how strikes are discussed in the UK, which is especially relevant in current affairs. As expected based on theory as well as claims from trade union leaders, the news and political environments are generally hostile towards strikes. This means that not only are citizens exposed to negative articles about the strikes in the newspapers, but this is also reflected in the political debates in the House of Commons. Strikes can be an important way for citizens to act and try to improve their pay and/or working conditions. A negative response from both the media and political actors influences the chances of strikes being successful. Thus, being aware of this negative bias in the news and politics is important to take into account when considering supporting or being against a strike, or even considering it as a potential option as a worker.

This negative environment is also what has potentially contributed to the introduction of the anti-strike Minimum Service Level Bill by the Conservative party. Bills like this along with negative news coverage could convince sceptics that the strikes are more disruptive than politically important as a means of effecting change. Additionally, considering how the Conservative party changed its policy stance on how to approach strikes, negative news may have influenced how MPs perceived the strikes. Therefore, the negative portrayal of the strikes in the news, combined with the outcomes of the VAR models that the negative newspaper articles are more influential on the speeches in parliament, suggests that how the strikes were covered by the news potentially impacted the political debate in parliament and as an extension also the new law being introduced. The way that the news portrays the strikes influences the understanding and interpretation of the strikes. If the news focused more on the positive aspects of the strikes, such as the awareness being raised for wages that are not reflective of the hard and demanding work, the political domain would potentially be more swayed towards addressing the root of the causes of the strikes, rather than the symptoms.

Besides empirical contributions and enriching the field of research on the relationship between media and politics, this thesis has also made multiple methodological contributions. The combination of deep-learning sentiment analysis and VAR modelling has not yet been used. What is more, the thesis was completed using only open-source NLP models available on Hugging Face instead of expensive services from companies such as Google or Amazon. Automating the process of sentiment analysis as well as creating the VAR models in Python made the entire analysis pipeline more robust and reproducible. All code has been made available on GitHub under a permissive license for future research.

7.1 Further research

Any further research into the relationship between media and politics should also introduce sentiment and political alignment into the research design, as this thesis has uncovered that these are very important factors to consider.

Further research into the relationship between the media and politics should build upon the work in this thesis to automate this type of analysis. This thesis has provided a way to use open-source NLP tools to analyse large amounts of text and use sentiment in VAR models. Other research could test this method further by experimenting with different deep-learning models.

As this thesis was the first to use only computational methods to consider the relationship between media and politics, it was limited to a high-profile case, the strikes in the UK. Future research could consider other topics to see if the results hold up for issues that are less publicised in the news. Additionally, more comparative research should be conducted to study this reciprocal relationship between news and politics in different countries.

References

Abercrombie, G. *et al.* (2019) 'Policy Preference Detection in Parliamentary Debate Motions', in *Proceedings of the 23rd Conference on Computational Natural Language Learning*, pp. 249–259. Available at: https://www.publicwhip.org.uk (Accessed: 6 November 2022).

Abercrombie, G. & Batista-Navarro, R. (2018) '"Aye" or "No"? Speech-level Sentiment Analysis of Hansard UK Parliamentary Debate Transcripts', in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, pp. 4173–4180.

Abercrombie, G. & Batista-Navarro, R. (2020a) 'ParlVote: A Corpus for Sentiment Analysis of Political Debates', in *Proceedings of the 12th Conference on Language Resources and Evaluation*, pp. 5073–5078. doi: 10.17632/czjfwgs9tm.1.

Abercrombie, G. & Batista-Navarro, R. (2020b) 'Sentiment and position-taking analysis of parliamentary debates: a systematic literature review', *Journal of Computational Social Science,* 3, pp. 245–270. doi: 10.1007/s42001-019-00060-w.

Abercrombie, G. & Batista-Navarro, R. T. (2022) 'Policy-focused Stance Detection in Parliamentary Debate Speeches', in *Northern European Journal of Language Technology, Volume 8*.

van Aelst, P. & Vliegenthart, R. (2014) 'Studying the Tango: An analysis of parliamentary questions and press coverage in the Netherlands', *Journalism Studies*, 15(4), pp. 392–410. doi: 10.1080/1461670X.2013.831228.

Austin, K. (2022) 'Train strikes: Rail workers to strike in run-up to Christmas - BBC News', *BBC,* 22 November. Available at: https://www.bbc.com/news/business-63715658 (Accessed: 26 April 2023).

Barberá, P. *et al.* (2019) 'Who Leads? Who Follows? Measuring Issue Attention and Agenda Setting by Legislators and the Mass Public Using Social Media Data', *American Political Science Review*, 113(4), pp. 883–901. doi: 10.1017/S0003055419000352.

Baumgartner, F. R. & Jones, B. D. (1993) *Agendas and instability in American politics.* University of Chicago Press (American politics & political economy).

BBC (2023) *As it happened: Teachers, doctors and Tube staff walk out in biggest strike day - BBC News*. Available at: https://www.bbc.com/news/live/uk-64957203 (Accessed: 16 March 2023).

Brandt, P. T. & Williams, J. T. (2007) *Multiple time series models.* SAGE (Quantitative applications in the social sciences: 148).

Chung, D. & Druckman, J. N. (2011) 'Identifying Frames in Political News', in Bucy, E. P. and Holbert, R. L. (eds) *Sourcebook for Political Communication Research: Methods, Measures, and Analytical Techniques*. New York, NY; London: Routledge, pp. 238–267. doi: 10.4324/9781315782713.

Cooper, G., Blumell, L. & Bunce, M. (2020) 'Beyond the "refugee crisis":

How the UK news media represent asylum seekers across national boundaries', *the International Communication Gazette*, 83(3), pp. 195–216. doi: 10.1177/1748048520913230.

Davies, M. & Nophakhun, R. (2018) 'Media "militant" tendencies; how strike action in the news press is discursively constructed as inherently violent', in Hart, C. and Kelsey, D. (eds) *Discourses of Disorder Riots, Strikes and Protests in the Media*. Edinburgh University Press, pp. 109–130. Available at: http://hdl.handle.net/10034/620941.

Dekker, R. & Scholten, P. (2017) 'Framing the Immigration Policy Agenda: A Qualitative Comparative Analysis of Media Effects on Dutch Immigration Policies', *The International Journal of Press/Politics*, 22(2), pp. 202–222. doi: 10.1177/1940161216688323.

Department for Business and Trade (2023) *Strikes (Minimum Services Levels) Bill 2023 - GOV.UK*. Available at: https://www.gov.uk/government/publications/strikes-minimum-services-levels-bill-2023 (Accessed: 8 March 2023).

Devlin, J. *et al.* (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1, pp. 4171–4186.

Dukes, R. & Kountouris, N. (2016) 'Pre-strike ballots, picketing and protest: Banning industrial action by the back door', *Industrial Law Journal*, 45(3), pp. 337–362. doi: 10.1093/indlaw/dww026.

Ellis, S. J. & Kitzinger, C. (2002) 'Denying Equality: An Analysis of Arguments against Lowering the Age of Consent for Sex between Men', *Journal of Community and Applied Social Psychology*, 12(3), pp. 167–180. doi: 10.1002/CASP.670.

Enevoldsen, K. C. & Hansen, L. (2017) 'Analysing Political Biases in Danish Newspapers Using Sentiment Analysis', *Journal of language works*, 2(2).

Entman, R. M. (1993) 'Framing: Toward Clarification of a Fractured Paradigm', *Journal of Communication*, 43(4), pp. 51–58. doi: 10.1111/j.1460-2466.1993.tb01304.x.

Erickson, C. L. & Mitchell, D. J. B. (1996) 'Information on strikes and union settlements: Patterns of coverage in a "newspaper of record"', *Industrial and Labor Relations Review*, 49(3), pp. 395–407. doi: 10.1177/001979399604900301.

Falck, F. *et al.* (2019) 'Measuring Proximity Between Newspapers and Political Parties: The Sentiment Political Compass', *Policy & Internet*, 12(3), pp. 367–399. doi: 10.1002/poi3.222.

Farrow, T. L. & O'Brien, A. J. (2005) 'Discourse analysis of newspaper coverage of the 2001/2002 Canterbury, New Zealand mental health nurses' strike', *International Journal of Mental Health Nursing*, 14(3), pp. 187–195. doi: 10.1111/J.1440-0979.2005.00380.X.

Finlayson, A. (2017) '"What is the point of parliamentary debate?" Deliberation, oratory, opposition and spectacle in the British House of Commons', *Redescriptions*, 20(1), pp. 11–31.

Flynn, F. J. (2000) 'No news is good news: The relationship between media attention and strike duration', *Industrial Relations*, 39(1), pp. 139–160. doi: 10.1111/0019-8676.00157.

Garland, R., Tambini, D. & Couldry, N. (2018) 'Has government been mediatized? A UK perspective', *Culture & Society*, 40(4), pp. 496–513. doi: 10.1177/0163443717713261.

Gillespie, J. (2021) 'Framing Strikes: A Case Study of Media Depictions of Two Teacher Strikes Teacher Strikes', in *NERA Conference Proceedings*. Available at: opencommons.uconn.edu/nera-2021/4 (Accessed: 9 March 2023).

Grijzenhout, S., Jijkoun, V. & Marx, M. (2010) 'Opinion mining in dutch hansards', in *Proceedings of the Workshop From Text to Political Positions, Free University of Amsterdam*. Citeseer, pp. 1–15.

Grijzenhout, S., Marx, M. & Jijkoun, V. (2014) 'Sentiment analysis in parliamentary proceedings', *From Text to Political Positions: Text analysis across disciplines*, 55, p. 117.

Grossman, E. (2022) 'Media and Policy Making in the Digital Age', *Annual Review of Political Science*, 25. doi: 10.1146/annurev-polisci-051120.

Hartmann, J. *et al.* (2023) 'More than a Feeling: Accuracy and Application of Sentiment Analysis', *International Journal of Research in Marketing*, 40(1), pp. 75–87. doi: 10.1016/J.IJRESMAR.2022.05.005.

Hertel-Fernandez, A., Naidu, S. & Reich, A. (2021) 'Schooled by Strikes? The Effects of Large-Scale Labor Unrest on Mass Attitudes toward the Labor Movement', *Perspectives on Politics*, 19(1), pp. 73–91. doi: 10.1017/S1537592720001279.

Hilton, S. *et al.* (2014) 'Implications for alcohol minimum unit pricing advocacy: What can we learn for public health from UK newsprint coverage of key claim-makers in the policy debate?', *Social Science & Medicine*, 102, pp. 157–164. doi: 10.1016/J.SOCSCIMED.2013.11.041.

Hughes, C. (2016) 'Its not easy (not) being green: Agenda dissonance of Green Party press relations and newspaper coverage', *European Journal of Communication*, 31. doi: 10.1177/0267323116669454.

Indulska, M., Hovorka, D. S. & Recker, J. (2012) 'Quantitative approaches to content analysis: Identifying conceptual drift across publication outlets', *European Journal of Information Systems*, 21(1), pp. 49–69. doi: 10.1057/ejis.2011.37.

Jansen, A. S. *et al.* (2019) 'Who Drives the Agenda: Media or Parties? A Seven-Country Comparison in the Run-Up to the 2014 European Parliament Elections', *The International Journal of Press/Politics*, 24(1), pp. 7–26. doi: 10.1177/1940161218805143.

JCHR (2023) *Strikes Bill fails to meet human rights obligations - JCHR -*

Committees - UK Parliament. Available at: https://committees.parliament.uk/committee/93/human-rights-joint-committee/news/186524/strikes-bill-fails-to-meet-human-rights-obligations-jchr/ (Accessed: 8 March 2023).

Jost, P. & Koehler, C. (2021) 'Who shapes the news? Analyzing journalists' and organizational interests as competing influences on biased coverage', *Journalism*, 22(2), pp. 484–500. doi: 10.1177/1464884918788270.

Juric, D., Hollink, L. & Houben, G. J. (2013) 'Discovering links between political debates and media', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7977 LNCS, pp. 367–375. doi: 10.1007/978-3-642-39200-9_30.

Juric, D., Houben, G.-J. & Hollink, L. (2012) 'Bringing parliamentary debates to the Semantic Web', in *CEUR Workshop Proceedings 902*, pp. 51–60. Available at: http://www.polimedia.nl/ (Accessed: 28 November 2022).

Kedia, A. & Rasu, M. (2020) *Hands-On Python Natural Explore tools and techniques to analyze and process text with*. Packt Publishing, Ltd.

Kleppe, M. *et al.* (2013) 'PoliMedia: Analysing media coverage of political debates by automatically generated links to radio & newspaper items', *CEUR Workshop Proceedings*, 1124, pp. 1–6.

Knowles, F. (2022) *Trade unions seen more negatively following 2022's summer strikes, YouGov*. Available at: https://yougov.co.uk/topics/politics/articles-reports/2022/12/02/trade-unions-seen-more-negatively-following-2022s- (Accessed: 9 March 2023).

Liu, Y. *et al.* (2019) 'RoBERTa: A Robustly Optimized BERT Pretraining Approach', in. Available at: http://arxiv.org/abs/1907.11692.

Lyddon, D. (2015) 'The changing pattern of UK strikes, 1964-2014', *Employee Relations*, 37(6), pp. 773–745. doi: 10.1108/ER-05-2015-0084.

Macaulay, T. (2016) 'How the papers covered the strike', *British Medical Journal*, 353. doi: 10.1136/bmj.i2506.

Matthes, J. & Kohring, M. (2008) 'The Content Analysis of Media Frames: Toward Improving Reliability and Validity', *Journal of Communication*, 58(2), pp. 258–279.

McCombs, M. E. (2013) *Setting the agenda: the mass media and public opinion*. Polity Press.

McCombs, M. F. & Shaw, D. L. (1972) 'The agenda-setting function of mass media', *Public Opinion Quarterly*, 36, pp. 176–187.

McCombs, M. & Valenzuela, S. (2020) *Setting the agenda: Mass media and public opinion*. John Wiley & Sons.

Milner, S. (2022) 'Another Winter of Discontent? Strikes in Britain', *Political Insight*, 13(3), pp. 40–43.

van Noije, L., Kleinnijenhuis, J. & Oegema, D. (2008) ‘Loss of Parliamentary Control Due to Mediatization and Europeanization: A Longitudinal and Cross-Sectional Analysis of Agenda Building in the United Kingdom and the Netherlands’, *Journal of Political Science,* 38(3), pp. 455–478. doi: 10.1017/S0007123408000239.

O’Neill, D. (2007) ‘From Hunky Heroes To Dangerous Dinosaurs: Journalism–Union Relations, News Access And Press Coverage In The 2002–3 British Fire Brigades Union Dispute’, *Journalism Studies*, 8(5), pp. 813–830. doi: 10.1080/14616700701504781.

Office for National Statistics (2023a) *Inflation and price indices*. Available at: https://www.ons.gov.uk/economy/inflationandpriceindices#:~:text=Consumer price inflation%2C UK%3A January 2023&text=The Consumer Prices Index including,from 9.2%25 in December 2022.

Office for National Statistics (2023b) *Labour disputes UK: total working days lost: all inds. & services (000’s)*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/timeseries/bbfw/lms.

Onyimadu, O. *et al.* (2014) ‘Towards Sentiment Analysis on Parliamentary Debates in Hansard’, in *Semantic Technology*, pp. 48–50. doi: 10.1007/978-3-319-06826-8_4.

PoliticsJOE (2022) *Mick Lynch goes in on the Daily Mail* . Available at: https://www.youtube.com/watch?v=Fnvv3EZHDnM&ab_channel=JOE (Accessed: 7 March 2023).

Radebe, M. J. (2006) *The Coverage of Industrial Action by the Mail & Guardian, 1999-2004*. University of the Witwatersrand, Johannesburg.

Rameshbhai, C. J. & Paulose, J. (2019) ‘Opinion mining on newspaper headlines using SVM and NLP’, *International Journal of Electrical and Computer Engineering (IJECE)*, 9(3), pp. 2152–2163. doi: 10.11591/ijece.v9i3.pp2152-2163.

Roggeband, C. & Vliegenthart, R. (2007) ‘Divergent framing: The public debate on migration in the Dutch parliament and media, 1995-2004’, *West European Politics*, 30(3), pp. 524–548. doi: 10.1080/01402380701276352.

Sawhney, R. *et al.* (2021) ‘GPolS: A Contextual Graph-Based Language Model for Analyzing Parliamentary Debates and Political Cohesion’, pp. 4847–4859. doi: 10.18653/v1/2020.coling-main.426.

Seabold, S. & Perktold, J. (2010) ‘Statsmodels: Econometric and Statistical Modeling with Python’, in *Proceedings of the 9th Python in Science Conference*, pp. 92–96. doi: 10.25080/majora-92bf1922-011.

Sellers, P. (2009) *Cycles of spin: Strategic communication in the US Congress*. Cambridge University Press.

Shapiro, A. H., Sudhof, M. & Wilson, D. J. (2022) ‘Measuring news sentiment’, *Journal of Econometrics*, 228(2), pp. 221–243. doi: https://doi.org/10.1016/j.jeconom.2020.07.053.

Smythe, P. (2022) ‘This summer’s strikes are already working – unions, set your sights even higher’, *The Guardian*, 21 July. Available at: https://www.theguardian.com/commentisfree/2022/jul/21/hot-strike-summer-britain-unions-spiralling-inflation (Accessed: 26 April 2023).

Soroka, S. N. (2002) *Agenda-setting dynamics in Canada*. UBC press.

Soroka, S. N. *et al.* (2002) ‘Issue Attributes And Agenda-Setting By Media, The Public, And Policymakers In Canada’, *International Journal of Public Opinion Research*, 14(3), pp. 264–285.

The Conservative Party (2019) *Conservative Party Manifesto 2019*. Available at: https://www.conservatives.com/our-plan/conservative-party-manifesto-2019 (Accessed: 8 March 2023).

Trades Union Congress (2017) *Trade Union Act 2016 - A TUC guide for union reps | TUC*. Available at: https://www.tuc.org.uk/research-analysis/reports/trade-union-act-2016-tuc-guide-union-reps (Accessed: 8 March 2023).

Transport for London (2021) ‘TfL warns of severe disruption due to planned RMT strikes’, 22 November. Available at: https://tfl.gov.uk/info-for/media/press-releases/2021/november/tfl-warns-of-severe-disruption-due-to-planned-rmt-strikes (Accessed: 7 March 2023).

Triggle, N. (2022) ‘December strike by NHS nurses is biggest in their history - BBC News’, *BBC*, 25 November. Available at: https://www.bbc.com/news/health-63746334 (Accessed: 26 April 2023).

Vliegenthart, R. *et al.* (2016) ‘Do the media set the parliamentary agenda? A comparative study in seven countries’, *European Journal of Political Research*, 55(2), pp. 283–301. doi: 10.1111/1475-6765.12134.

Vliegenthart, R. & Damstra, A. (2019) ‘Parliamentary Questions, Newspaper Coverage, and Consumer Confidence in Times of Crisis: A Cross-National Comparison’, *Political Communication*, 36(1), pp. 17–35. doi: 10.1080/10584609.2018.1478472.

Vliegenthart, R. & Montes, N. M. (2014) ‘How Political and Media System Characteristics Moderate Interactions between Newspapers and Parliaments: Economic Crisis Attention in Spain and the Netherlands’, *The International Journal of Press/Politics*, 19(3), pp. 318–339. doi: 10.1177/1940161214531335.

Walgrave, S. & van Aelst, P. (2006) ‘The contingency of the mass media’s political agenda setting power: Toward a preliminary theory’, *Journal of Communication*, 56(1), pp. 88–109. doi: 10.1111/j.1460-2466.2006.00005.x.

Walgrave, S., Soroka, S. & Nuytemans, M. (2008) ‘The Mass Media’s Political Agenda-Setting Power: A Longitudinal Analysis of Media, Parliament, and Government in Belgium (1993 to 2000)’, *Comparative Political Studies*, 41(6), pp. 814–836.

Walker, P. (2023) ‘Anti-strike bill: Shapps to get power to decide minimum service levels’, *The Guardian*, 10 January. Available at:

https://www.theguardian.com/uk-news/2023/jan/10/anti-strike-bill-unveiled-in-commons-but-no-detail-on-minimum-service-levels (Accessed: 8 March 2023).

Wankmüller, S. (2022) 'Introduction to Neural Transfer Learning With Transformers for Social Science Text Analysis', *Sociological Methods and Research*, 0(0), pp. 1–77. doi: 10.1177/00491241221134527.

Wells, R. & Caraher, M. (2014) 'UK print media coverage of the food bank phenomenon: from food welfare to food charity?', *British Food Journal*, 116(9), pp. 1425–1445. doi: 10.1108/BFJ-03-2014-0123.

Willis, R. (2017) 'Taming the Climate? Corpus analysis of politicians' speech on climate change', *Environmental Politics*, 26(2), pp. 212–231. doi: 10.1080/09644016.2016.1274504.

Wolf, T. *et al.* (2020) 'HuggingFace's Transformers: State-of-the-Art Natural Language Processing', in, pp. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6.

Yang, X. *et al.* (2023) 'Exploring the limits of chatgpt for query or aspect-based text summarization', *arXiv preprint arXiv:2302.08081*.

Appendices

Appendix A - VAR Model 1: Bivariate

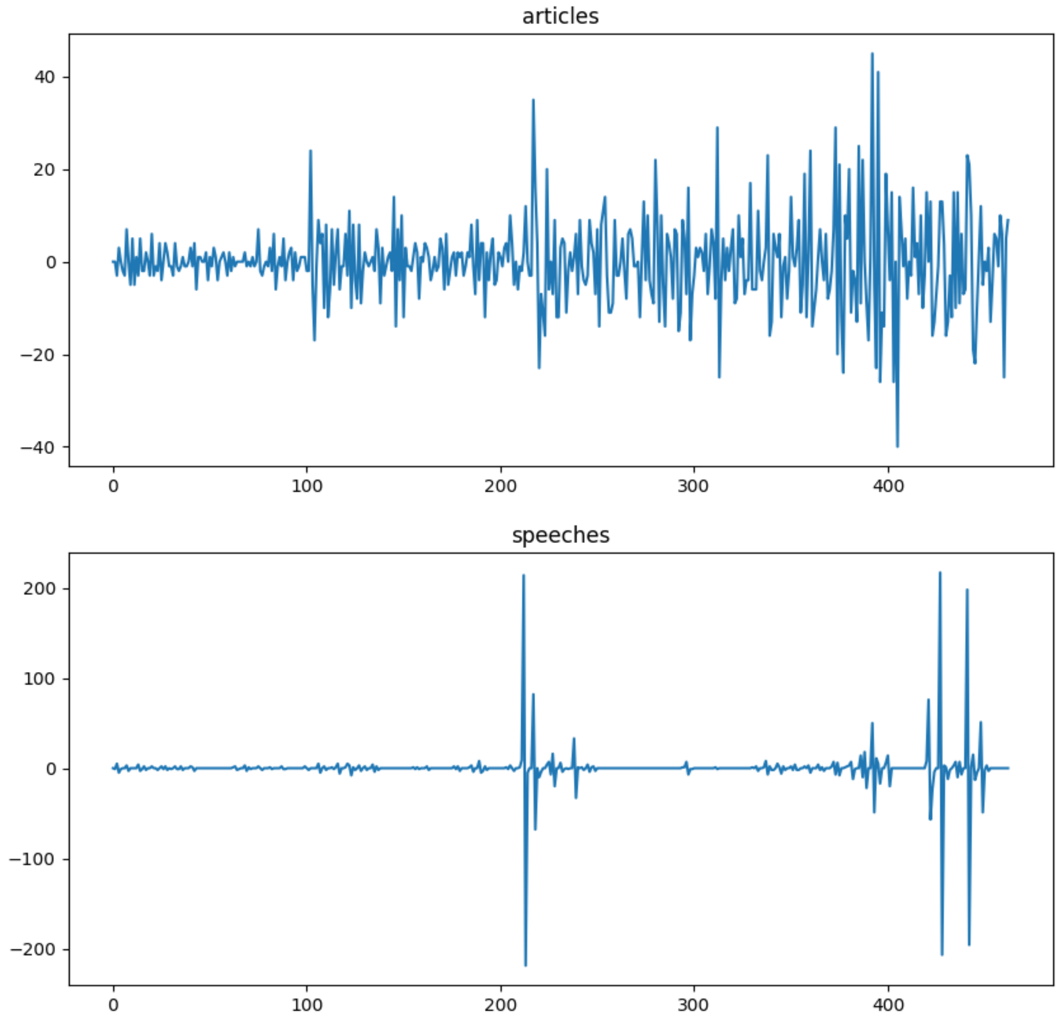

Figure 7: Articles and speeches timeseries after taking the first difference.

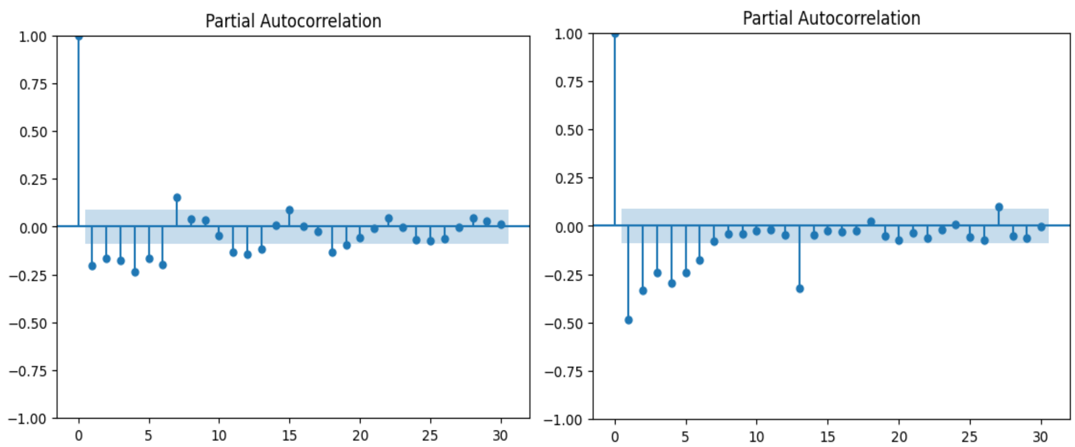

Figure 8: Partial autocorrelation plots for articles and speeches respectively.

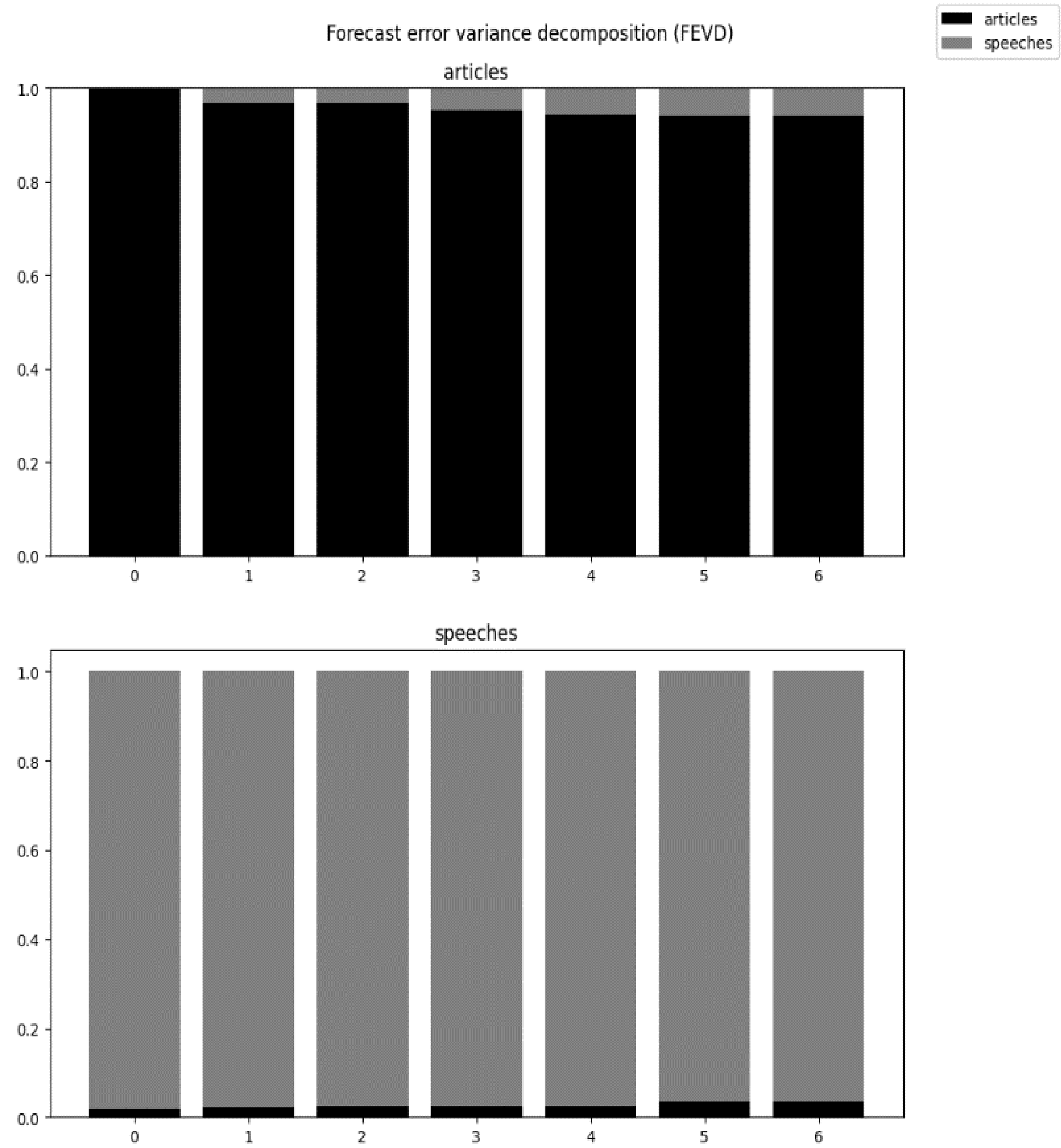

Figure 9: FEVD for articles and speeches

Appendix B - VAR Model 2: Sentiment

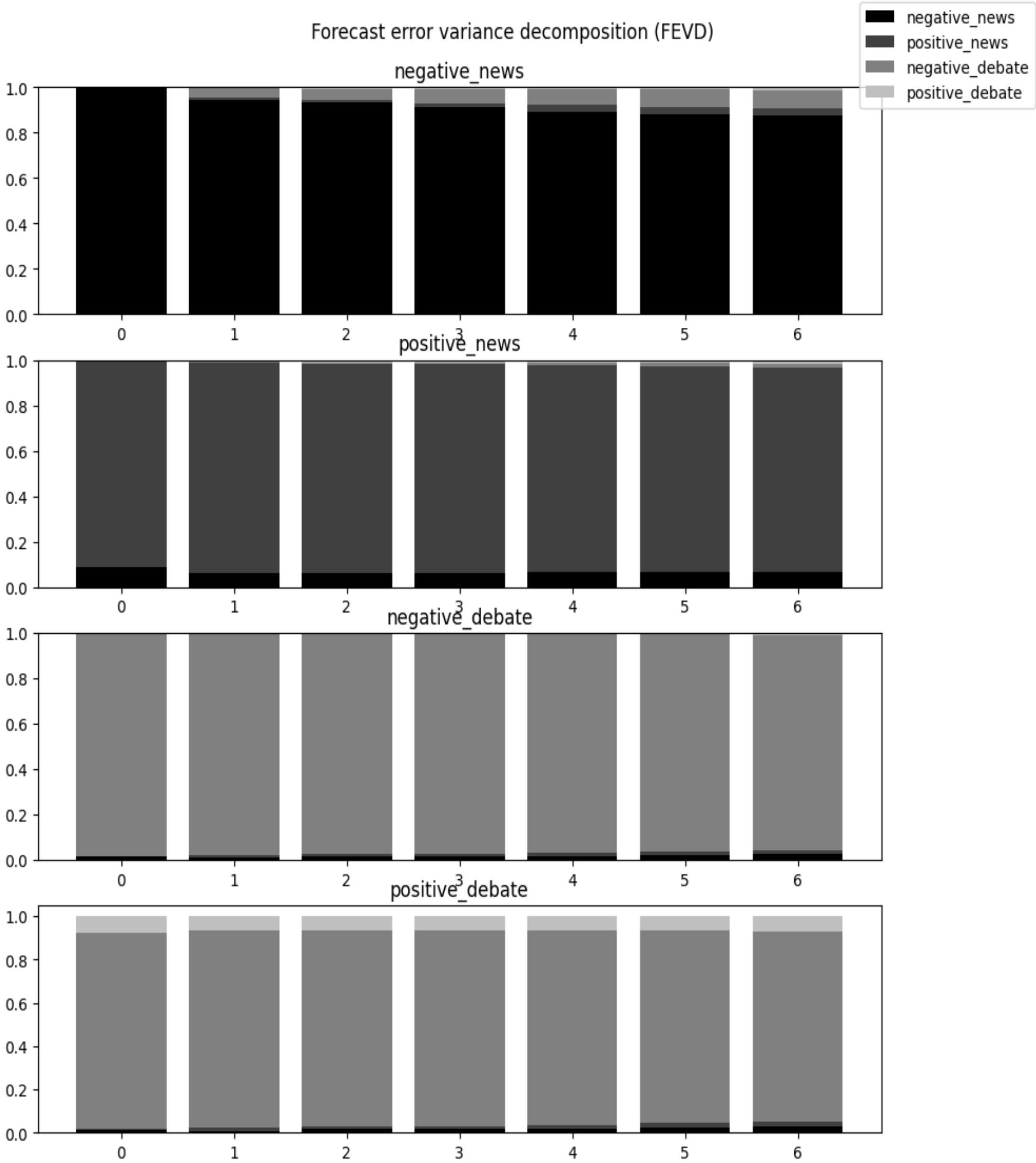

Figure 10: FEVD for articles and speeches by sentiment.

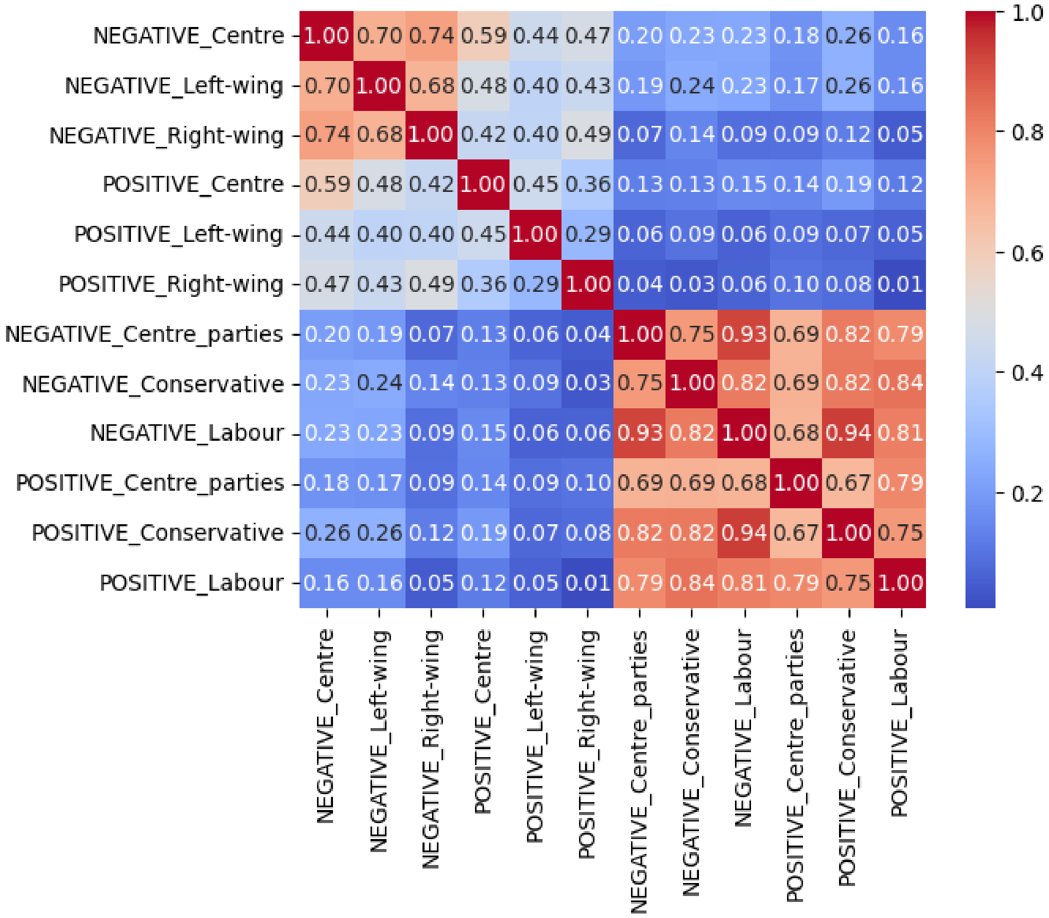

Figure 11: Correlation matrix for all the variables in VAR model 3.

Appendix C - VAR Model 3: Sentiment and Political Ideology

VAR Model 3 Part 1 (Speeches)

Articles	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
	Negative speeches – Centre parties					
Constant	0.00	0.08	0.02	0.98 n.s.		
Negative Centre $_{t-1}$	-0.08	0.04	-2.08	0.04 *	$_{(t-2)}$0.00 ***	2.81%
Negative Left-wing $_{t-1}$	0.16	0.06	2.76	0.01 **	0.01 **	0.81%
Negative Right-wing $_{t-3}$	0.10	0.05	2.10	0.04 *	0.00 ***	0.83%
Positive Right-wing $_{t-2}$	0.27	0.08	3.23	0.00 *	$_{(t-4)}$0.01 **	1.56%
	Positive Speeches – Centre parties					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	-0.00	0.02	-0.02	0.98 n.s.		
Negative Centre $_{t-1}$	-0.02	0.01	-2.06	0.04 *	0.00 **	0.77%
Negative Right-wing $_{t-3}$	0.03	0.01	2.18	0.03 *	0.00 ***	1.08%
Positive Right-wing $_{t-2}$	0.05	0.02	2.13	0.03 *	$_{(t-4)}$0.04 *	0.35%
Positive Left-wing $_{t-2}$	0.06	0.03	2.04	0.04 *	0.45 n.s.	0.23%
	Negative speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.01	0.21	0.05	0.96		
Negative Centre $_{t-3}$	-0.33	0.13	-2.55	0.01 *	0.00 ***	1.29%
Negative Left-wing $_{t-1}$	0.36	0.14	2.57	0.01 *	0.00 ***	0.49%
Negative Right-wing $_{t-4}$	0.38	0.11	3.28	0.00 **	0.00 ***	1.28%
Positive Right-wing $_{t-5}$	0.71	0.23	3.14	0.00 **	0.00 ***	0.62%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	-0.01	0.19	-0.04	0.97 n.s.		
Negative Centre $_{t-2}$	-0.22	0.11	-2.08	0.04 *	0.00 ***	3.66%
Negative Left-wing $_{t-4}$	-0.36	0.17	-2.09	0.04 *	$_{(t-3)}$0.00 **	3.43%
Negative Right-wing $_{t-3}$	0.33	0.10	3.17	0.00 **	0.00 ***	1.16%
Positive Right-wing $_{t-5}$	0.50	0.21	2.44	0.02 *	0.00 ***	1.18%
	Negative speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	-0.00	0.22	-0.01	0.99 n.s.		
Negative Centre $_{t-2}$	-0.35	0.13	-2.78	0.01 **	0.00 ***	3.34%
Negative Left-wing $_{t-1}$	0.31	0.15	2.07	0.04 *	0.01 **	0.57%
Negative Right-wing $_{t-3}$	0.34	0.12	2.82	0.01 **	0.00 ***	1.12%
Positive Right-wing $_{t-2}$	0.50	0.22	2.32	0.02 *	$_{(t-4)}$.01 *	1.58%
	Positive speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.004	0.07	0.06	0.95 n.s.		
Negative Centre $_{t-3}$	-0.14	0.04	-3.14	0.00 **	0.00 ***	0.71%
Negative Left-wing $_{t-1}$	0.11	0.05	2.26	0.02 *	0.00 **	0.40%
Negative Right-wing $_{t-4}$	0.10	0.04	2.55	0.01 *	0.00 ***	0.81%
Positive Right-wing $_{t-5}$	0.16	0.08	2.08	0.04 *	0.00 ***	1.18%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

VAR Model 3 Part 2 (Articles)

Speeches	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
	Negative articles – Centre newspapers					
Constant	0.03	0.12	0.27	0.79 n.s.		
Negative Centre parties $_{t-1}$	0.48	0.21	2.23	0.03 *	0.04 *	2.80%
Negative Conservative $_{t-4}$	0.21	0.10	1.99	0.05 *	0.01 **	2.25%
Positive Centre parties $_{t-1}$	2.14	0.49	4.38	0.00 ***	0.01 *	1.56%
Positive Labour $_{t-1}$	-0.92	0.24	-3.80	0.00 ***	$_{(t-4)}$0.00 ***	2.64%
Positive Conservative $_{t-2}$	-0.31	0.13	-2.44	0.02 *	$_{(t-3)}$0.03 *	1.35%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.03	0.10	0.24	0.81 n.s.		
Negative Centre parties $_{t-7}$	-0.53	0.19	-2.79	0.00 **	$_{(t-3)}$0.03 *	2.62%
Negative Labour $_{t-7}$	0.24	0.11	2.22	0.03 *	$_{(t-3)}$0.01 *	0.48%
Positive Labour $_{t-3}$	-0.52	0.26	-1.98	0.05 *	0.08 n.s.	0.33%
Positive Conservative $_{t-1}$	-0.28	0.09	-3.08	0.00 **	$_{(t-2)}$0.04 *	1.12%
	Negative articles – Right-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.01	0.13	0.08	0.93 n.s.		
Negative Centre $_{t-2}$	1.47	0.69	2.13	0.03 *	0.00 **	1.48%
Negative Conservative $_{t-5}$	0.25	0.11	2.21	0.03 *	$_{(t-3)}$0.00 **	2.82%
	Positive articles – Right-wing Newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.01	0.06	0.19	0.85 n.s.		
Negative Centre $_{t-4}$	-0.35	0.17	-2.02	0.04 *	0.02 *	3.66%
Positive Labour $_{t-3}$	0.22	0.11	2.07	0.04 *	0.00 **	0.50%
Negative Conservative $_{t-1}$	0.11	0.05	2.27	0.02 *	0.03 *	0.57%

Positive Conservative $_{t-1}$	-0.13	0.06	-2.29	0.02 *	0.00 **	0.44%
Negative articles – Left-wing Newspapers						
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.00	0.08	0.01	0.99 n.s.		
Negative Centre parties $_{t-2}$	0.40	0.19	2.06	0.04 *	0.00 **	0.98%
Negative Conservative $_{t-3}$	0.15	0.07	2.36	0.02 *	0.00 ***	1.41%
Positive Labour $_{t-5}$	-0.40	0.20	-2.03	0.04 *	0.01 *	1.59%
Positive Conservative $_{t-5}$	-0.23	0.09	-2.51	0.01 *	0.00 **	0.21%
Positive articles – Left-wing Newspapers						
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.01	0.05	0.23	0.82 n.s.		
Negative Labour $_{t-1}$	0.11	0.06	1.97	0.05 *	0.34 n.s.	0.29%
Negative Conservative $_{t-3}$	0.11	0.04	2.66	0.00 **	0.63 n.s.	0.34%
Positive Speeches Labour $_{t-3}$	-0.38	0.13	-2.93	0.00 **	0.35 n.s.	0.57%
Positive Conservative $_{t-2}$	-0.12	0.05	-2.34	0.02 *	0.73 n.s.	1.25%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Appendix D - VAR Model 4: Sentiment and Political Ideology Summer 2022

VAR Model 4 (Summer 2022)

	Negative articles – Right-wing Newspapers					
Speeches	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	-0.03	0.58	-0.06	0.96 n.s.		
Positive Centre parties $_{t-2}$	9.28	4.15	2.24	0.03 *	0.22 n.s.	1.88%
Positive Speeches Labour $_{t-2}$	-7.54	2.60	-2.90	0.00**	$_{(t-1)}$0.05 *	6.11%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.03	0.16	0.18	0.86 n.s.		
Negative Conservative $_{t-1}$	0.26	0.08	3.22	0.00 **	0.54 n.s.	3.31%
Positive Speeches Conservative $_{t-2}$	-0.91	0.42	-2.16	0.03 *	0.70 n.s.	4.97%
	Negative speeches – Centre parties					
Articles	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.02	0.25	0.07	0.95 n.s.		
Negative Left-wing $_{t-1}$	0.42	0.18	2.37	0.02 *	0.17 n.s.	9.01%
	Positive speeches – Centre parties					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.00	0.11	0.03	0.98 n.s.		
Negative Left-wing $_{t-1}$	0.17	0.08	2.10	0.40 *	0.26 n.s.	10.28%
	Negative speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.07	0.88	0.08	0.94 n.s.		
Negative Left-wing $_{t-1}$	1.73	0.64	2.69	0.01 **	0.24 n.s.	8.74%
	Positive speeches – Labour party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.02	0.37	0.05	0.96 n.s.		
Negative Left-wing $_{t-1}$	0.67	0.27	2.51	0.01 *	0.35 n.s.	8.32%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.05	0.82	0.06	0.95 n.s.		

Negative Left-wing $_{t-1}$	1.48	0.59	2.49	0.01 *	0.16 n.s.	9.09%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Appendix E - VAR Model 5: Sentiment and Political Ideology Winter 2022-3

VAR Model 5 (Winter 2022-3)

	Negative articles – Right-wing Newspapers					
Speeches	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	-0.19	0.51	-0.37	0.71 n.s.		
Negative Centre parties $_{t-2}$	2.58	1.23	2.10	0.04 *	0.12 n.s.	3.36%
Negative Labour $_{t-2}$	-1.50	0.66	-2.26	0.02 *	0.10 n.s.	4.40%
Positive Conservative $_{t-2}$	1.11	0.55	2.03	0.04 *	0.13 n.s.	1.43%
	Positive articles – Centre newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.04	0.41	0.09	0.93 n.s.		
Positive Centre parties $_{t-1}$	6.49	2.09	3.10	0.00 **	0.38 n.s.	11.56%
Positive Labour $_{t-1}$	-1.82	0.89	-2.04	0.04 *	$_{(t-3)}$0.04 *	3.80%
Positive Conservative $_{t-5}$	-0.65	0.30	-2.17	0.03 *	0.49 n.s.	6.94%
	Positive articles – Left-wing newspapers					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.01	0.13	0.07	0.95 n.s.		
Negative Conservative $_{t-1}$	-0.39	0.13	-3.05	0.00 **	0.71 n.s.	0.20%
Positive Labour $_{t-1}$	-0.58	0.28	-2.05	0.04 *	0.99 n.s.	2.90%
	Negative speeches – Centre parties					
Articles	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.29	0.39	0.75	0.46 n.s.		
Negative Centre $_{t-1}$	-1.02	0.27	-3.77	0.00 ***	$_{(t-2)}$0.00 **	4.79%
Negative Left-wing $_{t-2}$	1.07	0.32	3.34	0.00 **	0.06 n.s.	17.82%
Negative Right-wing $_{t-3}$	0.94	0.34	2.74	0.01 **	0.00 **	3.37%
Positive Centre $_{t-4}$	1.38	0.52	2.67	0.01 **	0.11 n.s.	8.26%
Positive Right-wing $_{t-2}$	1.02	0.38	2.65	0.01 **	0.44 n.s.	37.14%
	Positive speeches – Centre parties					

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.03	0.07	0.41	0.68 n.s.		
Negative Centre $_{t-2}$	-0.15	0.05	-2.97	0.00 **	0.00 ****	3.11%
Negative Left-wing $_{t-4}$	-0.19	0.08	-2.42	0.02 *	0.01 **	9.47%
Positive Left-wing $_{t-1}$	0.28	0.13	2.09	0.04 *	0.86 n.s.	6.93%
Positive Right-wing $_{t-2}$	0.19	0.07	2.67	0.01 **	0.08 n.s.	28.69%

Negative speeches – Labour party

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.74	0.94	0.78	0.44 n.s.		
Negative Centre $_{t-2}$	-2.42	0.66	-3.67	0.00 **	0.01 **	8.51%
Negative Left-wing $_{t-1}$	2.09	0.78	2.69	0.01 **	0.14 n.s.	10.26%
Negative Right-wing $_{t-2}$	1.66	0.78	2.14	0.03 *	0.05 *	1.14%
Positive Centre $_{t-4}$	3.17	1.26	2.51	0.01 *	0.16 n.s.	5.25%
Positive Right-wing $_{t-1}$	-1.79	0.83	-2.16	0.03 *	0.50 n.s.	17.64%

Positive speeches – Labour party

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.19	0.24	0.78	0.44 n.s.		
Negative Centre $_{t-1}$	-0.60	0.17	-3.51	0.00 ***	$_{(t-2)}$0.00 **	7.35%
Negative Left-wing $_{t-1}$	0.44	0.20	2.21	0.03 *	0.03 *	8.70%
Negative Right-wing $_{t-4}$	0.40	0.15	2.61	0.01 **	0.00 **	3.31%
Positive Right-wing $_{t-1}$	-0.53	0.22	-2.48	0.01 *	0.56 n.s.	19.12%

Negative speeches – Conservative party

	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.29	0.57	0.52	0.60 n.s.		
Negative Centre $_{t-2}$	-0.87	0.40	-2.18	0.03 *	0.01 **	7.06%
Negative Left-wing $_{t-1}$	1.84	0.47	3.93	0.00 ***	0.05 *	13.85%
Negative Right-wing $_{t-4}$	0.71	0.36	1.99	0.05 *	0.02 *	10.15%

Positive Articles Right-$_{t-2}$	-1.60	0.75	-2.13	0.03 *	0.71 n.s.	20.29%
	Positive speeches – Conservative party					
	coefficient	SE	t-stat	p-value	Granger Causality	FEVD
Constant	0.57	0.86	0.67	0.51 n.s.		
Negative Left-wing$_{t-2}$	-2.77	1.24	-2.23	0.03 *	0.17 n.s.	16.37%
Negative Right-wing$_{t-3}$	1.61	0.76	2.12	0.03 *	0.01 *	2.26%

***p < 0.001 **p < 0.01 *p<0.05 n.s. not significant
SE = Standard Error t-stat = t-statistic
FEVD = forecast error variance decomposition

Appendix F – Overview of All VAR Models

Comparison of VAR models

	Model 1	Model 2	Model 3	Model 4	Model 5
Description	Frequency only	Frequency and sentiment	Frequency, sentiment and political alignment	Same as model 3 but during summer 2022	Same as model 3 but during winter 2022-3
Maximum Lags	6	6	7	3	5
AIC	10.07	12.45	9.95	7.01	11.22
Notable results affecting speeches	Significant Granger tests, 2.4% FEVD	Similar FEVD: 2.4% (neg) and 2.9% (pos)	Conservative party highest FEVD 3.7% (pos)	Negative left-wing very significant	Positive right-wing very significant
Notable results affecting articles	Significant Granger tests, 3.2% FEVD	Negative highest FEVD: 4.5%	Centre speeches most influential with FEVD	Only two papers with significant predictors	Three papers with predictors, all influenced by Labour
Notable results overall	Relationship significant both directions	Negative also predicts positive	Negative more often significant	FEVD values go up, fewer significant Granger tests	FEVD values go up, Granger less often significant

pos = positive neg = negative
FEVD = forecast error variance decomposition

Table 12: Comparison of VAR models