Measuring Polarization using Newspaper Data

Validation and Application

Master Thesis Behavioural Economics

Erasmus School of Economics | Erasmus University Rotterdam

Abstract

This thesis develops a new daily metric to measure the level of polarization in the American public, by exploiting the discourse used in political news articles. Due to frequency discrepancies, formal comparison of the invented method to previous measurements of polarization is not possible. The metric shows that, in contrary to popular belief, the trend in polarization seems to decrease till 2009 after which the trend evolves in a cyclical pattern. The volatility of the metric declines, indicating an increasing stability over time. Having daily estimates for polarization, a vast array of potential uses to determine the causes and effects of changing polarization is unlocked. Demonstrated as an application is the "election cycle effect", seeking to determine if the year after a presidential election yields different changes in polarization compared to the other years. Formal testing shows that for the total sample, this is the case.

Author
T.A. Kloosterboer[*]

Student Number
483863tk

Supervisor
Dr. V. Spinu

First Reader
Dr. G.D. Granic

Date
October 22, 2018

[*] Contact – T.A.Kloosterboer@student.eur.nl or Tim.Kloosterboer@hotmail.com
Erasmus University Rotterdam | Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

Table of contents

1 Introduction

On August 12, 2017, the governor of Virginia declared a state of emergency in Charlottesville, VA as members of far-right movements clashed with counter protestors over a plan to remove the statue of a Confederate general from a city park. After the event turned violent, among the casualties were a 19-year old woman, 2 state troopers and well over 40 non-fatal injured persons[2]. This is not the sole example of clashes between ideological opposites in the United States. From the University of Maryland stabbing in 2017[3] to clashes in Portland (OR) in 2018[4], recent evidence that shows violent encounters between movements of contrarian ideologies seems ample. Hence, popular literature is quick to label the rising polarization as a threat to society.

In contrary to popular belief, scientific evidence is still mixed whether the polarization is increasing over the past decades or whether it is a misconception based on the wrong interpretation of available evidence. Basically, studies can be divided in two schools of thought, distinguished by different definitions of polarization. One claims polarization has increased in the electorate as well as among the political elites causing Democratic and Republican voters to be now sharply divided in their policy preferences (Abramowitz, 2010; Abramowitz and Saunders, 2008). The second school contends that the electorate is merely sorted better and that party supporters remain much more moderate than party elites, rejecting the claim of rising polarization (Fiorina et al., 2005; Fiorina and Abrams, 2008; Levendusky, 2009). What both schools have in common, is the fact that they base their findings on low-frequency survey data, data is collected solely every two or four years. As data is collected over a long period, long term changes are noticed but the available data prohibits studying more subtle changes, effects and causes.

Current study tries to address this data issue by employing a new method to measure polarization. By exploiting the language utilized in newspapers, a daily metric is developed that tracks the level of polarization in the American society. Newspaper data is used as it follows the ideology of the audience. For example, the earlier discussed "Unite the Right rally" in Charlottesville, VA is described differently by newspapers that appeal to a different public, causing conservative media to use a different discourse compared to more liberal media outlets. Scientific literature backs the relationship between the ideology of an audience and the language used by its media source. Evidence shows the media influences the beliefs of the audience (Andrews and Caren, 2010; Weaver, 2007; Corbett and Durfee, 2004) and also influences the perceived importance attached to each particular topic (McCombs, 2002). At the same time, different studies

[2] *"Three dead after white nationalist rally in Charlottesville"* published on August 12, 2017 by PBS. Extracted from https://www.pbs.org/newshour/amp/nation/state-emergency-charlottesville-va-fights-erupt-white-nationalist-rally on October 10, 2018.

[3] *"Former University of Maryland student indicted on hate crime charge in stabbing death"* published on October 18, 2017 by CNN. Extracted from https://edition.cnn.com/2017/10/17/us/university-of-maryland-student-hate-crime-charge on October 10, 2018.

[4] *"Clash in Portland points to City's Deeper Racial Divide"* published on August 5, 2018 in the New York Times.

show that media outlets follow the beliefs of the audience, both in political- (Entman, 1989) as well as in other contexts (Gans, 1979).

Hence, the polarization metric capitalizes on the fact that the language in newspapers follows the beliefs of its public and will shift together with changes in their ideology. Using the *Wordscores* algorithm (Laver, Benoit and Garry, 2003) – which is discussed in Section 3 – every political news article is rated with an ideology score based on the words used in that respective article. Subsequently, the polarization metric is determined by employing a moving variance on the article scores. Primarily, this study tries to validate and to review the characteristics of this metric. However, by having daily polarization levels, it yields opportunities to study relationships between other concepts and their correlation with the polarization levels in the US electorate. Formerly, this used to be impossible as polarization metric were determined on an (at least) bi-yearly level. By example, this study showcases an instance of such an application. I study whether the presidential election cycle has an influence on the levels of polarization in the American public. That is, if specific years in the cycle induce higher levels of polarization. Summarizing, this study tries to add to the existing evidence on the trends in polarization by applying a new method to measure this polarization via the use of the employed language in political news articles. Further, after validating the metric, an example of an application based on the benefits of the new, daily metric is illustrated.

The remainder of this study is laid out as follows: Section 2 will start by reviewing whether newspaper data is a credible source for measuring the ideology its audience, followed by studying the existing evidence on polarization in the US and identifies the theoretical evidence on our studied application. Section 3 describes how the data is collected, adjusted to arrive at the polarization metric and concludes with the methodology on how the hypotheses are tested. Section 4 shows both the descriptive statistics as well as the results of our analysis while the final chapter, Section 5 discusses the results of this study and offers possibilities for future research.

2 Literature review

2.1 The use of newspaper data in scientific studies

As outlined in the introduction, this study intends to use a new method to measure the polarization levels in the American society, via the language used in newspapers. However, the use of newspaper data to explain both collective actions and thoughts is not new. Data collected via traditional media sources such as newspapers and television broadcasts have been used in various studies. For example, Woolley (2000) identified public signals about monetary policy, Galambos (1975) studied long-term changes in public opinion and there have been multiple studies on media influence on the willingness to protest and consequential, collective violence (fe. Myers, 2000). Nevertheless, at its origin, newspaper content is not created for conducting research nor is it intended to be a perfect representation of reality. As a result, critical questions should and have been asked about the validity of using popular media sources for scientific purposes (Barranco and Wisler, 1999). Barranco and Wisler (1999) determine that newspapers are not a transparent channel of information, and that systematic selection- and description bias affect the types of data available in newspaper. However, Earl et al. (2004) note in defense of using newspaper data, all social scientists face the issue that they usually work with imperfect data. Hence, Earl and others (2004) argue that despite its issues, researchers may still effectively use newspaper data and that the data does not deviate substantially from accepted standards of quality. For example, the 78 % rate for interest in political rallies based on newspaper data by Oliver and Myers (1999) reflects the rates obtained in national surveys. While equally imperfect, survey data is widely used in scientific research. Therefore, whether to use newspaper data depends primarily on the purpose of the data but if approached with caution, newspaper data remains a useful data source (Baumgarten and Grauel, 2009).

Following Baumgarten and Grauel (2009), the characteristics of the media source need to be assessed to reflect whether newspaper data may be used for the goals of this study. As stated by Baumgarten and Grauel (2009), studies with a research purpose not solely interested in the individual communication strategies but with a more focused concern on the structures and dynamics of the comprehensive public discourse may have a valid data source in media data. Further, the direction and strength of the biases depends mainly on the adopted perspective (Baur and Lahusen, 2005). From the perspective of current study, the level of bias in topic selection is part of the variable of interest. Hence, fully correcting it would decrease the observed effect and is undesired. Therefore, I determine that the language used in newspapers may be adopted for the purposes of this study.

2.2 The reflection of ideological views by the discourse in newspapers

After establishing that newspaper data may be used as a valid data source in this study, it is yet to ascertain that it accurately proxies the public ideology. However, it can be assumed that journalists modify stories to their liking. This causes news articles to be subject to external influences (Ortiz et al., 2005). A standard practice in business would be to adjust your product or service to the wishes of the customer. Newspaper companies are likely to follow this rationale, since the circulation numbers of a newspaper are its prime source of income, and so it cannot afford to displease its audience. Bennett (1988) confirms this by showing that journalists modify publications in such a way that they match the perceived interest and social impact of the audience. This concurs with nationwide evidence on climate change. Research has shown that the media have mainly conveyed an image of certainty and scientific knowledge on climate change in Germany (Weingart et al., 2000) while emphasizing uncertainty of this environmental issue in the

US (Antilla, 2005). This follows a survey conducted by Gallup[5] in 2010, where substantially more Germans perceive climate change as a threat compare to Americans. Additional evidence by Gans (1979) suggests that American journalists adhere to typical American values such as individualism and share aversion against phenomena such as social disorder and ideological excesses. Entman (1989) provides evidence in a more political context, concluding that the political messages of newspapers are significantly associated with the political attitudes of their readers. So, it can be assumed that journalists adjust the reality to publish articles that are more reflective of the opinion of their consumers, irrespective of possible minor inconsistencies with the truth.

Different studies provide evidence that there is a reverse effect. That is, the publications of the media influence the audience. By drawing attention to specific issues and claims, the news media can shape the public opinion (Andrews and Caren, 2010). Specifically, by making use of framing techniques, the media can impact the public on what they should care about (Weaver, 2007). Corbett and Durfee (2004) confirm this using solely newspaper data, as both public perception and public attitudes are significantly effected by the articles published in the written press. Also, readers and viewers attach different levels of importance to topics based on the emphasis placed on it (McCombs, 2002). Concluding, this evidence shows that newspapers have a role to play in shaping the public's concern for an issue as well as the perceived relevance of that specific issue.

Concluding. the literature provides evidence in two directions: An influence of media content on the beliefs of the audience as well an effect in the opposite way, that the informational needs of the audience influence the content published by the media. This study exploits this implied correlation between media content and the beliefs of the audience. However, as a basis of the developed polarization metric, it remains important to confirm whether this implication is valid. Hence, I hypothesize:

> *H_1: The textual bodies of political news articles reflect the ideological views of its audience.*

2.3 Trends in the US polarization levels

The evidence on polarization in the US is focused on two distinct areas, elite polarization and mass polarization. Elite polarization encompasses polarization among the political elite, party members and elected officials. It rises when these elites converge their stances on policy positions with members of the same party while diverging their policy positions with members of opposite parties. Evidence shows that in recent decades, these political elites have become more ideological dissociated and they are more likely to take on extreme positions regarding political matters

[5] Report by Gallup, "Fewer Americans, Europeans View Global Warming as a Threat" published on April 20, 2011 on news.gallup.com and retrieved on July 5, 2018.

(McCarty et al., 2006; Theriault, 2008). This rise in elite polarization has significant consequences for partisan identities, voting behaviour and the policy preferences of the mass public (Hetherington, 2001; Iyengar et al., 2012; Layman and Carsey, 2002). As a result, elite polarization reshaped American politics into less consensus and more extremes. However, this implied consequence of the rising elite polarization on the electorate – which is of interest in respect to this study – remains tentative. What is apparent though, is the fact that elite polarization has led to increased recognition of party differences and a heightened sense that the outcome of elections matters (Jacobsen, 2000). Clearly, this indicates that elite polarization may gradually increase mass polarization as the party differences do rise. However, Pinsky (2006) shows that this process can well be disturbed by changing behaviour of the political elites, for example, the Republican party nominating a colored presidential candidate.

In contrary to elite polarization, the debate in the scientific literature on mass polarization is deeply divided. This divide is mainly caused by varying definitions of polarization, as both ideological alignment and ideological divergence are used as definitions to proxy for mass polarization (Lelkes, 2016). Primarily, two sides in identifying polarization can be distinguished. The one side believes America is in the middle of a culture war, together with increasingly high levels of polarization. This side – generally represented by Alan Abramowitz and Kyle Saunders – claims polarization should be defined as ideological alignment, referring to the concept of matching ideology and the level to which attitudes become more consistent. Abramowitz and colleagues show that the percentage of voters that places them self in the middle of the ideological spectrum has decreased significantly over the years (Abramowitz and Saunders, 2008; Abramowitz, 2010). On the other side, Fiorina and colleagues define polarization as ideological divergence or the degree to which the distribution of public ideology has moved apart. They conclude that claims of a culture war are exaggerated and that there is no conclusive evidence of rising polarization. They claim that solely America's elected officials and activists are sharply divided into two different ideological corners, but that the silent majority is primarily moderate and does remain centrally located in the ideological spectrum (Fiorina et al., 2005; Fiorina and Abrams, 2008, Levendusky, 2009). Most likely, both these schools find evidence in support of their stance due to the difference in research approach and construction of the polarization concept. Therefore, consensus remains absent and neither a clear definition nor a clear measurement of polarization is established.

This study follows the school of Fiorina, using ideological divergence as interpretation of the polarization concept. Hence, it is hypothesized that there is no evidence of rising polarization in the American public.

H_2: The level of polarization is not rising in the American public.

2.4 Stability in the polarization metric

As established, it is common practice by newspapers to shift publications into the ideological direction its audience is most affiliated with. While this phenomenon is not new, improvements in technologies to govern the clientele have made it possible to track the news consumption of the audience in more quantifiable details (Napoli, 2010). This causes newspapers to have a better view of preferences and consequently, they have been able to adhere more effectively to the wishes of the public (Vu, 2014). Additionally, besides consuming news, audiences now also contribute heavily to the spreading and creation of the news (Baek et al., 2011). Meraz (2011) confirms, "traditional media's singular, one-way power over news creation and dissemination is a past phenomenon". Using web-based technologies such as social media citizens can connect easily online to discuss and promote the content of the news. As this increases the reach and income, producing articles that appeal to the audience yields increasingly more benefits (Meraz, 2011). Additionally, from a reader's viewpoint, the internet and other technological changes such as social media have provided the audience with limitless options for selecting the source of their news. This increased buyer power further shaped the editors of media outlets to be more sensitive to both positive and negative signals from their audience.

The improved ability and need to follow consumer preferences does not indicate that a newspaper would never publish anything contrary to the beliefs of its audience. However, it does give them the ability to adjust its messages in the preferred direction and to track what kind of news yields the most profitable results. So, due to improved monitoring newspapers are better at judging and adjusting to the needs of customers. This is increasingly important due to the vast number of alternatives for reading news that became available with the rise of the internet (Dimmick et al., 2004). Hence, as the metric exploits a single source in a longitudinal study, it is expected that the polarization metric will get more stable over time. Fewer extreme changes will decrease the differential and thus, clarify the trend. This causes me to to hypothesize:

H_3: The polarization metric stabilizes over time.

2.5 Applying the metric: The election cycle effect

This study introduces a measure to obtain daily polarization scores. Therefore, it yields opportunities for studying the impact and causes of polarization at a more detailed level. Previously, the absence of high frequency data prohibited developing structured studies on the polarization trajectory, causes and effects. This paper showcases an application of the high-frequency metric by investigating whether the presidential election cycle influences the polarization level. That is, whether the amount of years to the next presidential election influences the changes in polarization visible in the metric.

In other segments of scientific research, there is evidence of such a presidential election cycle effect. For example, US stock prices tend to closely follow the 4-year presidential election cycle. In general, stock prices fall steeply the first six months after a president has been elected, reach a low in the second year, after which they rise towards a peak in either the third or the fourth year of the election cycle (Wong and McAleer, 2009). Further, Zhao, Liano and Hardin (2004) argue that during the final two years of the presidency the economy is stimulated to enhance the re-election chances of the party of the incumbent president using fiscal, economical and administrative policies. The interpretation by Zhao, Liano and Hardin (2004) may be extended to the purpose of this study. First, the start of a new government primes citizens to consider the introduced policies in prospective terms. Regardless of whether a challenger promises "change" or a re-elect vows to "take care of unfinished business", the start of a new term more or less resets the views of the public. As the presidency advances, the electorate begins to judge the doings of the elect and starts to move towards more stricter views on the implemented policies. This either strict agreement or disagreement to introduced policies increases the polarization (Singer and Carlin, 2013).

Another factor influencing the polarization is the level of political engagement in the electorate (Hetherington, 2008). Presidential elections are typically times of increased political engagement, where levels of political activity are higher compared to periods between elections (Vitak et al., 2011). This increase in political activity gets strengthened by the use of internet and social media, especially by people that used to be less involved in politics (Baumgartner and Morris, 2010; Johnson and Perlmutter, 2010). As increased political activity can usually be translated in stronger views on ideological subjects, it is apparent that the further the election cycle proceeds towards the election day the polarization will increase as well. Thus, based on reasoning originating from the economic presidential cycle effect as well increased political engagement, I hypothesize:

H_4: The later in the presidential election cycle, the higher the level of polarization.

3 Research method

3.1 Data collection

The political news articles are extracted from the *LexisNexis Academic* database, which is an online repository of journalistic content from all over the world. Within the database, the source is set to *The New York Times*. As a newspaper among the highest in circulation in the United States, and with a reputation of thoroughness and credibility it is preferred over other sources. Ideally the dataset would comprise of all the news articles published in the US. However, due to feasibility constraints, this is not possible. This study determines the trajectory of the polarization on a longitudinal basis rather than studying absolute values in a cross-section. Hence, the absolute starting values are less of an issue and using solely the NYT data will not prohibit our study from making valid estimations on the trend.

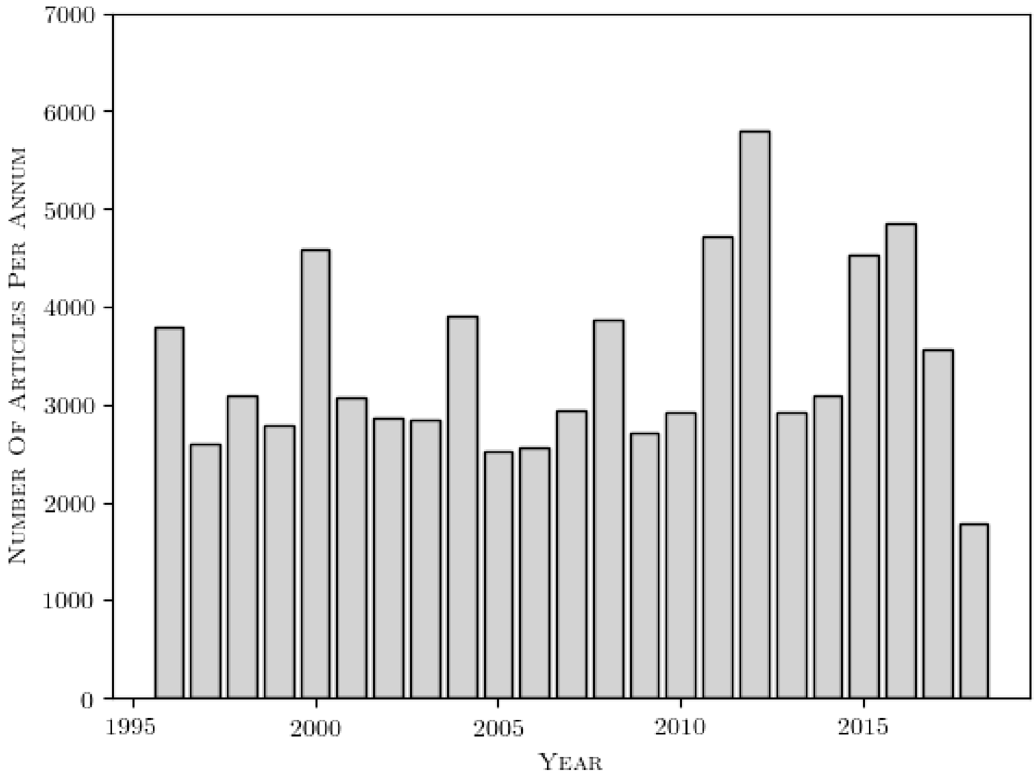
LexisNexis' internal taxonomy is used to identify and select articles that contain US political news. Specifically, all articles are selected that meet both of the following two tags: *'Election and politics'* (a sub tag of 'Government and Public Administration') and *'US'* (a sub tag of 'North America'). Selected are all articles that follow these criteria and are published in the time period between January 1, 1996 and May 31, 2018.

This selection yields over 250.000 articles. However, articles such as "*On Visit to Congo, Albright Praises the New Leader*" [6] and "*Howard Fast, 88, Best-Selling Novelist, Dies*" [7] do meet the requirements set but are still not very relatable to the goal of this study. Hence, excluded from our final dataset are all articles that are not published in either the '*US; politics*' or the '*National Desk*' sections of *The New York Times*. This gives the most uniform dataset in terms of the subject, i.e. internal US matters with a political emphasis. Further, articles with less than 40 words are removed from our dataset. In that case, it is expected that either the text of the article solely comprises of the caption beneath an image or that the database contains an error. This selection results in a dataset of a total of 78.314 articles, relatively equally distributed over the timespan studied. A visual representation of the distribution of articles over the years is provided in Figure 1. Major difference in the distribution can be explained by a rise in articles published in election years (1996; 2000; 2004; 2008; 2012 and 2016) and to a lesser extent, in pre-election years (1999; 2003; 2007; 2011 and 2015), which is logical considering the rise in interest in internal politics in these years. Consequently, a higher publication rate of politically orientated articles by *The New York Times* can be seen.

[6] *"On Visit to Congo, Albright Praises the New Leader"* published on December 13, 1997 in *The New York Times* regarding a Congo rebel army seizing power earlier in 1997 and a US governmental visit to the country.
[7] *"Howard Fast, 88, Best-Selling Novelist, Dies"* published on March 13, 2003 in *The New York Times* regarding the death of best-selling historical fiction book writer Howard Fast.

Figure 1. Distribution of articles per year

The distribution of the articles used in this study divided per year. All articles are published in *The New York Times*. Note that the timespan of the study ranges to May 2018, hence the number of articles is lower in 2018 compared to previous years.



Within each article, removed are words that comprise of solely capitals. This removes the location of publication of the article and abbreviations which could lead to spurious correlations and artificial changes in scoring the article. Also, punctuation and numbers are removed. Finally, as standard in quantitative text analysis, words are converted to lowercase only.

3.2 Determining the ideological score

For calculating the ideological scores of each article used is a supervised learning approach. In supervised learning the machine is fed a number of pre-scored cases, the reference data. The algorithm learns from the reference data how a difference in frequency of a single word influences the score of an article. For example, the word '*illegal'* is more frequent is conservative-minded documents and so will receive a higher, more right, score from the algorithm. Analyzing each individual news article in our dataset is done by applying the *Wordscores* algorithm created by Laver, Benoit and Garry (2003). Based on this standalone algorithm, policy positions are extracted by interpreting text not as discourse to be understood but as data in the form of words. This breaks with previous methods that determine ideological scores in text, such as the Comparative Manifestos Project (CMP), in a sense that text does not need to be interpreted by a coder or by a computer program applying a pre-coded dictionary. It decreases the influence of human judgment as well as the necessity for human intervention. *Wordscores* calculates – given a set of reference documents – an ideological position of a body of text about which nothing is known using word frequency matrices.

The first step in determining the ideological score of each news article in the sample is to determine the reference documents against which the articles are paralleled. As clearly, the usefulness of the *Wordscores* metric depends heavily on the ability to select the appropriate reference texts (Slapin and Proksch, 2008). Used are the Democratic- and Republican Party Platforms as extremes on the left-right scale[8]. Even though it is debatable that the platforms are the extremes on an ideological scale, they do provide guidance and clarity. As our dataset comprises of bodies of text from 1996 to 2018, using a platform from a single election would unrightfully assumes that the political lexicon is constant over time (Slapin and Proksch, 2008). Hence, used as reference texts are the party platforms of both the Republican- and the Democratic party from all election years within the dataset (1996, 2000, 2004, 2008, 2012 and 2016). I follow Budge and Pennings (2006) by assigning the ideological scores the Comparative Manifesto Project[9] (CMP) collected to each party platform. Note that the CMP scores are low for Democrats and high for Republicans. This causes the words used in Democratic platforms in years where the CMP rated the platform as more left on the ideological scale to also have a more leftish impact on the article scores as rated by the Wordscores algorithm. The used reference scores for each party platform for both the Democratic party and the Republican party can be found in Appendix A.

Using the party manifestos and their respective scores, a word score is determined for each individual word occurring in the reference texts. Let F_{wt} be the relative frequency of word w in training document t. The probability that reference text t is read given that word w can be seen is $P(t|w) = F_{wt} / \sum F_{wt}$. As A_t is set as the overall score of the reference texts determined by the CMP, each individual word can be scored according to $S_w = \sum [P(t|w) \cdot A_t]$. So, each total word score is the sum of the multiplication of the article scores and the relative occurrence of the respective word in each article. For example, observed is that the word 'risk' is used 25 times per 10,000 words in reference text *Z* and 75 times per 10,000 words in reference text *Y*. That simply induces that if the word 'risk' is read, the posterior probability that the reference text *Z* is read is 0.75. If text *Z* its reference score is 10 while text *Y* has a reference score of -10, that yields a word score for 'risk' of 0.25(-10) + 0.75(10) = +5.0. Increasing the number of reference texts – as is the case in this study – does not alter the technique for calculating the word scores but solely increases the

[8] Both the Democratic- and Republican Party Platforms are collected from "The American Presidency Project" retrieved at http://www.presidency.ucsb.edu/ on 1st of August 2018.
[9] Values obtained as determined by the Comparative Manifestos Project (CMP), retrieved from https://www.manifesto-project.wzb.eu on 3th of August 2018.

computational complexity. As simply put, the score of a word is a weighted sum of word contributions to each article weighted by article score.

Table 1 contains the words that, based on the reference texts, have the most left and right impact on the article scores. Within the top words on the left side of the ideological spectrum, important liberal values are represented: Education (preschool; educate; scientists), climate change (temperature), racial equality (Muslims; Latinos). The words with heaviest load to the right show strong interaction with important factors in conservative spheres: National security (safely; Iraq; armies; espionage), traditional values (safely; pride; conventional) and law enforcement and immigration control (trial; attorney; sponsored). This shows that despite using party platforms as ideological extremes in reference texts is not optimal, the *Wordscores* algorithm is still able to distinguish between left and right, score-wise, on the ideological scale.

Table 1. Words with most impact as determined by *Wordscores*

This table contains the top 20 words to both the left, liberal side as well as the right, conservative scale after applying the *Wordscores* algorithm on the party platforms of the Democrat- and Republican party platforms.

Left		Right	
Preschool	Childcare	Safely	Governor
Wellbeing	LGBT	Pride	Armies
Lowcost	Muslims	Strength	Shortages
Temperature	Nurture	Trial	Contagious
Scientists	Inspire	Landowners	Cartel
Antitrust	Discriminate	Normal	Unfunded
Disproportionate	Latinos	Attorney	Confrontation
Educate	Marijuana	Conventional	Rhetoric
Inequality	Tolerance	Iraq	Sponsored
Resilient	Arbitrary	Reconciliation	Espionage

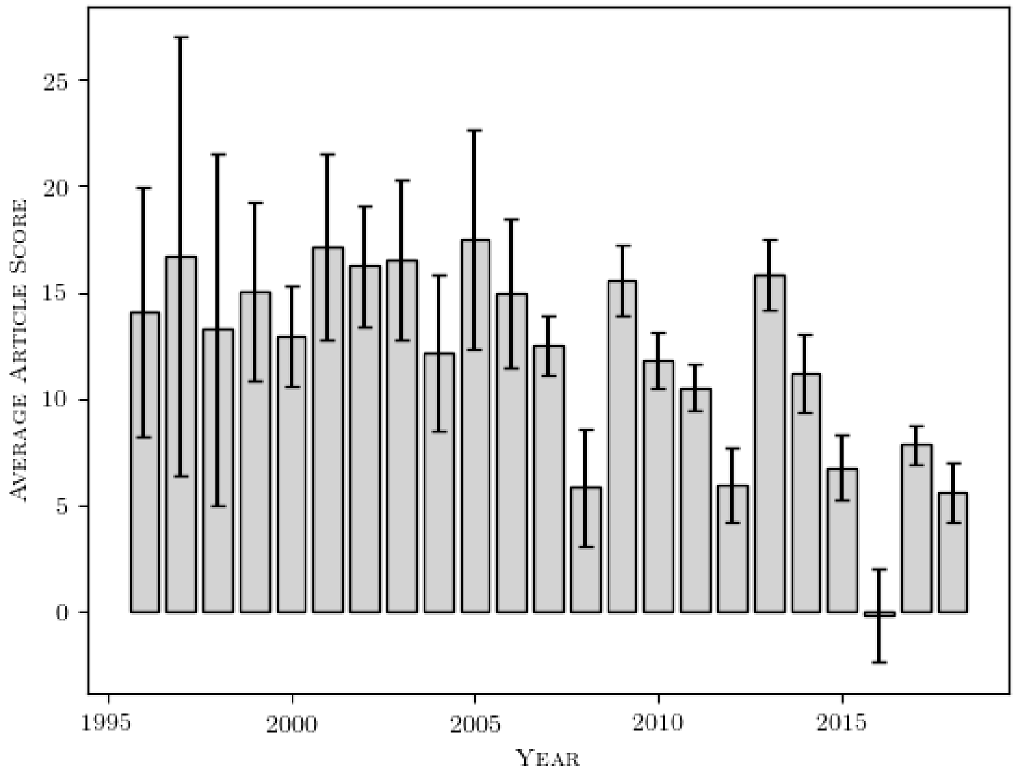
The next step is to compute the scores of each news article. F_{wn} is the relative frequency of each word w in the news article n. Causing the score of each news article to be $S_n = \sum (F_{wn} \cdot S_w)$. Simplified, the relative frequency of each word multiplied by its word score and summed across all words in a news articles gives the article score.

Finally, the scores are rescaled. As in any text, the most frequent words are stop words (‘the’, ‘and’, ‘that’, etc.). As these words will have similar relative frequencies across both the Democratic and Republican reference texts, all the document scores will be bunched around the middle. Hence, following Laver, Benoit and Garry (2003), I correct for this ‘bunching’ of test scores as follows: $S_v^* = (S_n - S_{\bar{n}})(\sigma_t/\sigma_n) + S_{\bar{n}}$, where S_n is the uncorrected score for news article n, $S_{\bar{n}}$ the average score of all the news articles, σ_t is the standard deviation of the reference scores, and σ_v is the standard deviation of all the news article scores. This formula transforms the articles scores by equalling the standard deviation in the article scores to the reference scores. Martin and Vanberg (2008) propose an alternative re-scaling formula to control for the fact that the transformed scores are not on the same scale as the original reference documents, but Benoit and Laver (2008) reply that their original formula is preferred in case of many test documents and few reference documents, as is the case in this research.

Appendix B contains the summary statistics for the article ideological scores. Besides the total sample, the scores are split on the respective year until the articles are collected. A visual representation of the yearly averages and standard errors are shown in Figure 2 below. A lower score indicates a more left, liberal article as it has closer resemblance to the Democrats' political lexicon. Surprisingly, the average ideological score seems to be decreasing over time, which could indicate that the audience of the NYT is changing its political preference to a more liberal stance. It is also a possibility that the Democratic reference texts, comprises of more subjects that became newsworthy in the latter part of the timespan. For example, traditionally important subjects for liberals such as global warming and gender equality seem to be increasingly relevant to the public, and hence the newspapers. Even though average scores do not influence the polarization, more focused topic selection indicates less spread in newspaper content and consequently, a lower polarization score. Whether this is completely due to the changing public, is questionable. Also noticeable are the substantially lower left-right scores on election years. All of 2000, 2004, 2008, 2012 and 2016 score low, on average, compared to the years around them. This indicates that the left-wing lexicon becomes more prominent close to the general election. Even though this does not directly yield evidence towards our fourth hypothesis that the election cycle influences the levels of polarization, it does indicate that the publications of political news articles are influenced by some sort of cyclical influence. Further, the standard error bars show that the volatility in the scoring of the news articles seems decreasing. The 63% confidence interval as indicated by the cap at the end of the error bars is substantially smaller in the latter years of the sample compared to the first few years of the sample. This is a first indication the polarization scores might be lower in the latter part of the sample.

Figure 2. Average article scores per year

This average article scores are obtained by applying the *Wordscores* algorithm on the extracted articled published in *The New York Times* from January 1996 to May 2018. The bars show the average article scores per year and the error bars indicate the standard errors.



3.3 Polarization metric

As discussed, this study supports the view that polarization can be seen in the divergence of ideological views. Usually in studies following this definition of polarization, the public is denoted polarized if the distribution of responses has bimodal characteristics (Fiorina et al., 2005; Fiorina and Abrams, 2008). A difference with current study is however, that in these studies responses are measured on a Likert 7-points liberal-conservative ideological scale based on survey responses. This study determines the ideology on an interval scale and hence simply using the bimodality metric would result in a loss of spread in the outcomes. Hence, a different computation is used to arrive at the polarization scores.

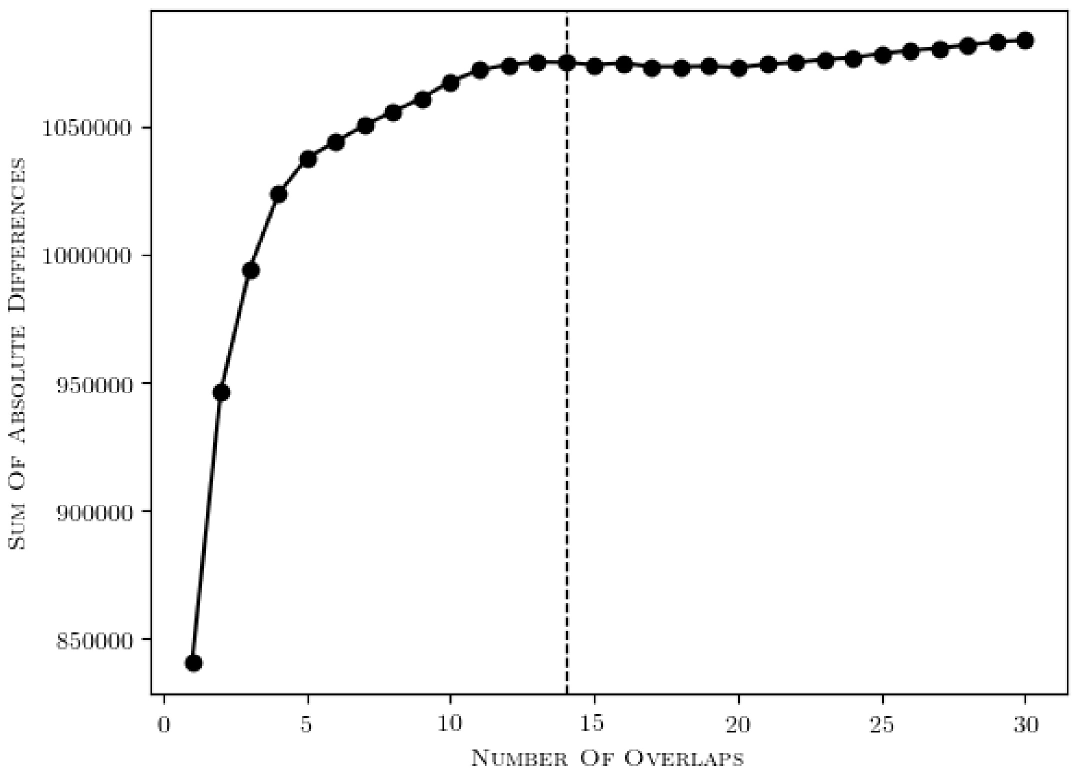
Given polarization's prominence in contemporary political discourse, there is little guidance and agreement in defining and measuring the concept. DiMaggio, Evans and Bryson (1996) acknowledge four all-embracing dimensions of polarization: *Dispersion*, *Bimodality*, *Constraint*, and *Consolidation.* Since the nature of the data in current study (a single distribution versus multiple distributions) rules out the usage of both the *Constraint* and the *Consolidation* dimensions and yet determined is that using the B*imodality* principle defers the distinctive advantages of the spread in the data. Hence, this paper follows previous studies (e.g., Baldassarri and Bearman, 2007; Evans, 2003; Alwan and Tufis, 2016) in identifying the *dispersion* dimension to measure polarization. Following these studies, variance is used as a metric for *dispersion* and consequently, as a proxy for polarization. This still bears close resemblance to the measurement of bimodality as employed by the school of Fiorina and others and hence our hypothesis does not need to be adjusted. As the two factors which comprises into the bimodality coefficient (BC), skewness and excess kurtosis, yield an effect in the identical direction in both the BC as well as in the variance. That is, an increasing skewness increases both the BC and the variance while an increasing excess kurtosis decreases both the BC and the variance.

However, using plain daily variance to measure polarization yields several disadvantages. First, polarization levels can fluctuate massively due to the nature of the data: news articles. A sudden event can swiftly change the majority of the news articles published into a particular direction and hence the accompanying left-right scores with it. Exemplified, around a school shooting, the news will be dominated by both articles in favor of gun control and opposers. As much it is likely that this will enhance the level of polarization in the society, the changes might be exaggerated in both strength and swiftness. This is strengthened by the fact that certain days in our studied time period – especially on Saturday and Sunday – the number of articles is low. This makes these days even more susceptible to sudden events. This can disturb the natural flow of the polarization metric and create extreme, and thus unwanted, changes over time.

I account for these disadvantages by applying a moving window to calculate the variance. Instead of calculating a daily variance based on the observations solely from that day, used are observations from 14 days before our day of interest as well as 14 days after our day of interest. Used are both lagged- and leading-days in articles scores to adjust for the delayed publication factor of news in a newspaper. Also, by using a symmetrical window, the moving variance will not be ahead of its time. Even though the effect will probably be minor in this case, preferred is to avoid this bias all together (Hyndman and Athanasopoulos, 2018). The amount of 14 days before and after the day of interest is chosen for multiple reasons. First, including more than 14 days does not add any more smoothing advantages. I arrive at this conclusion by calculating the sum of the absolute differences between the scores of the different window sizes and the scores just calculating it by using the single day observations. The total summed differences are plotted against the amount of days included in the moving window. In Figure 3 can be seen that after the moving window reaches the size of 14 days (both before and after the day of interest) the sum of the absolute difference starts to flatten out. Hence, I conclude that increasing the window size does not produce much more smoothing benefits. Second, incorporating the nature of the data, incorporating a moving window removes any concerns of low levels of published articles on holidays such as Christmas, control for Saturday and Sunday, and control for sudden events with a major impact on the political news, such as the September 11 attacks in 2001 or the Iraq invasion in 2003. Finally, it also captures the fact that polarization is investigated as a process which changes over time. As a large timespan is studied, losing resolution to obtain a smoother polarization metric is acceptable. A disadvantage of the overlapping of observations is that it creates an autocorrelated error term and consequently, causes any ordinary least squares parameter estimation to be inefficient. This leads to biased standard error estimates in performed hypothesis tests (Hansen and Hodrick, 1980).

Figure 3. Sum of absolute differences on number of lags added

Optimal number of overlapping days to include in the moving average calculation to determine the polarization score. The sum of absolute differences is calculated by taking the total difference of adding no overlaps and the respective overlaps both lagged and leading visible on de horizontal axis.



Summarizing, I measure polarization using a moving daily variance with the window sizes of -14 and +14 days on the ideology scores obtained by applying the *Wordscores* algorithm on the political news articles published in *The New York Times*.

3.4 Methodology
Validation of the metric

As this study employs a new method to determine the levels of political polarization it is vital to assess its validity. Based on the existing literature hypothesized is that using the textual bodies of political news articles reflects the ideological view of the American public. Comparing it to preceding methods of measuring polarization may confirm whether it captures, and possibly adds to the explanatory power by these former approaches. Commonly, political polarization is measured by using data collected in surveys, notably from the American National Election Studies[10] (ANES) and the General Social Surveys[11] (GSS). These surveys measure the views of the American citizens on a broad range of societal issues. In both the ANES and the GSS, a 7-point Likert scale asks respondents their view on a liberal-conservative scale ranging from extremely liberal to extremely conservative. Using the answers on this question, per study changing ways of conducting calculations generally yield the polarization metric. An issue with the collected data from ANES and GSS is that the frequency of the conducted surveys is lower compared to the polarization metric obtained by using news articles. The GSS is collected every two years, while the ANES is also collected every two years except for 2006, 2010 and 2014. As current metric introduces daily data, this immediately highlights the advantage of our method but complicates validating the metric substantially.

In order to be able to compare the developed metric with previous metrics and hence to test the first hypothesis, all article scores are assigned to their respective year. This limits the frequency discrepancy compared to survey methods of calculating polarization. So, in order to test this hypothesis, the newspaper metric is calculated by taking a regular variance of all *Wordscore*s article scores within a single year. That is, the variance of article scores ranging from January 1st, 2017 to December 31st, 2017 will be the 2017 polarization score. According to Lelkes (2016), Freeman and Dale (2013) and following Fiorina et al. (2008) their interpretation of polarization, the polarization of the ANES and GSS data is determined by calculating the bimodality coefficient on the survey responses (BC):

1) $$Bimodality\ Coefficient\ (BC) = \frac{m_3^2 + 1}{m_4 + 3 \cdot \frac{(n-1)^2}{(n-2)(n-3)}}$$

Where m_3 refers to the skewness of the distribution, m_4 refers to the excess kurtosis and n is the annual sample size. To provide a visualization of the movement from all three time-series, the calculated polarization scores for the ANES, GSS and current metric are plotted over time in Section 4.2. Finally, to formally assess whether the different metrics are significantly correlated over time, pairwise correlations are calculated and assessed on their significance. As both the ANES and GSS data do not have a sufficient frequency, more advanced methods of comparing the methods are infeasible and would yield unreliable results.

[10] ANES data collected from: *“https://electionstudies.org/data-center/”* on August 28, 2018.
[11] GSS data collected from: *“http://gss.norc.org/get-the-data”* on August 28, 2018.

Trends in the level of polarization

In contrary to the previous hypothesis, the developed daily polarization metric is used. Visual evidence of the trend and the descriptive statistics are discussed in Section 4.1. To formally test the trend of polarization over time, I use Newey-West regression to adjust for the serial-correlation in the standard errors. These heteroskedasticity- and autocorrelation-consistent standard errors are specified to include a maximum of 14 lags, as the periodicity of the calculated metric is 14 days back. This causes the Newey-West standard errors to be robust to both arbitrary autocorrelation as well as arbitrary heteroskedasticity. To control for possible exogeneous influences on changes in the polarization levels, added are control variables.

I add the daily stock price for the "New York Times Company", the media company that publishes *The New York Times* and is traded at the New York Stock Exchange (NYSE). This controls for price pressure from the public on its publications. As a rational manager should desire to achieve a high stock price in order to maximize firm value, they should be susceptible to price pressure. Further, a dummy variable for the partisan identity of the president is added. It might be the case that a particular party induces more polarized articles than the other, and if that is the case, it needs to be corrected for. Additionally, a categorical variable is added for the executive editor of the NYT. Within the period of interest of this study, the position of executive editor changed 5 times. It might be possible that with the change of executive editor, the emphasis on parts of the news changes or that certain directors are more susceptible to sales pressure, and hence influences our polarization metric. Also, controlling for potential seasonality, a categorical variable for the month of the year is added. Regressions are both run on the condensed equation solely consisting of the independent variable representing date as well as the full equation, which looks as follows:

2) $$Polarization_t = \alpha + \beta_1 \cdot Date_t + \beta_2 \cdot Stock\ Price\ NYT_t + \beta_3 \cdot Party\ President + \beta_4 \cdot Exec.Editor + \beta_5 \cdot Month_t + \varepsilon_t$$

Volatility of the polarization metric

Also, assessed it whether the stability of the polarization metric changes over time. In order to examine whether this is the case, I run a Newey-West regression on the absolute first differences in the polarization metric. Again, Newey-West regressions are run to correct for the autocorrelation in the error term. Even though the polarization metric is first-differenced, the correlograms still show evidence of autocorrelation, and hence Newey-West standard errors are used. Also, as there is no need to make forecasts based on the estimates, I decided against the use of more complicated models such as the ARCH or GARCH models. The Newey-West regressions are run on both the condensed formula solely including the time variable as well as on the extended formula including all of the control variables. The extended formula looks as follows:

3) $\Delta\, Polarization_t = \alpha + \beta_1 \cdot Date + \beta_2 \cdot Stock\ Price\ NYT_t + \beta_3 \cdot Party\ President + \beta_4 \cdot Exec.Editor + \varepsilon_t$

where delta polarization is the absolute first-difference between the polarization on date *t* and date *t-1* and the independent variables are identical to the ones used in equation 2).

As it is hypothesized that the polarization metric becomes more stable over time, a significant decrease in the coefficient for the differences in polarization against time should be visible. If this is the case, it can be concluded that the changes in the polarization metric become less volatile and hence, are more likely to represent the true trend. Especially since big daily variations in the polarization metric are unlikely to happen frequently. Since even after terrorist attacks or the financial crises, the whole of the population will see shifts in ideology which softens or even cancels out the effect on the polarization metric.

Application: The effect of the election cycle

To examine the effect each year in the election cycle has on the level of polarization, I follow Allvine and O'Neil (1980) in calculating a yearly percentual change in polarization. This percentual change is calculated between the polarization values on the respective starting- and ending dates in Table 2. That is, for the 2016 election on November 8, polarization scores on November 9th, 2016 of 200 and November 8th, 2017 of 220 would yield a percentual change of +10%. So, considering the boundaries of the dataset, the four groups show percentual changes for the 1996, 2000, 2004, 2008, 2012 and 2016 elections[12].

Table 2. Observations to include in each election cycle year

This table lists which observative relative to the election are used to determine the annual change in the polarization metric. For each election, the change is calculated by dividing the polarization level at the ending day by the polarization level at the starting day and subtracting 1.

Year in Election Cycle	Group	Starting day rel. to election day	Ending day rel. to election day
-2	0	-366	-729
-1	0	-1	-365
+1	1	1	365
+2	0	366	729

Percentual changes are preferred over absolute changes due to the fact the starting values are not equal over time. Using percentages gives direct insight into the true scale of differences rather than being dependent on the outright values. This would disproportionately overweight the first cycles as the starting values are higher compared to the latter cycles.

[12] Bar -2 for the 1996 election, -1 for the 1996 election and +2 for the 2016 election. Due to data constraints these changes are either based on partial observations or missing entirely.

First, the values are aggregated and plotted in a bar chart to obtain visual evidence whether the change in the level of polarization might be dependent on the year in the election cycle. Subsequently, run is a Mann-Whitney U test to test for the differences in the distributions between the first year after the election and the other years in the cycle (Mann and Whitney, 1947). That is, on the differences in changes between group 0 and 1 in Table 2. Even though the nature of the polarization metric – using overlapping windows – violates the independence assumption of the Mann-Whitney U test, this minor dependence is assessed as no factor to discard the use of the test. The Mann-Whitney U test will show whether there are significant differences between the distributions of the changes in the level of polarization between both groups.

4 Results

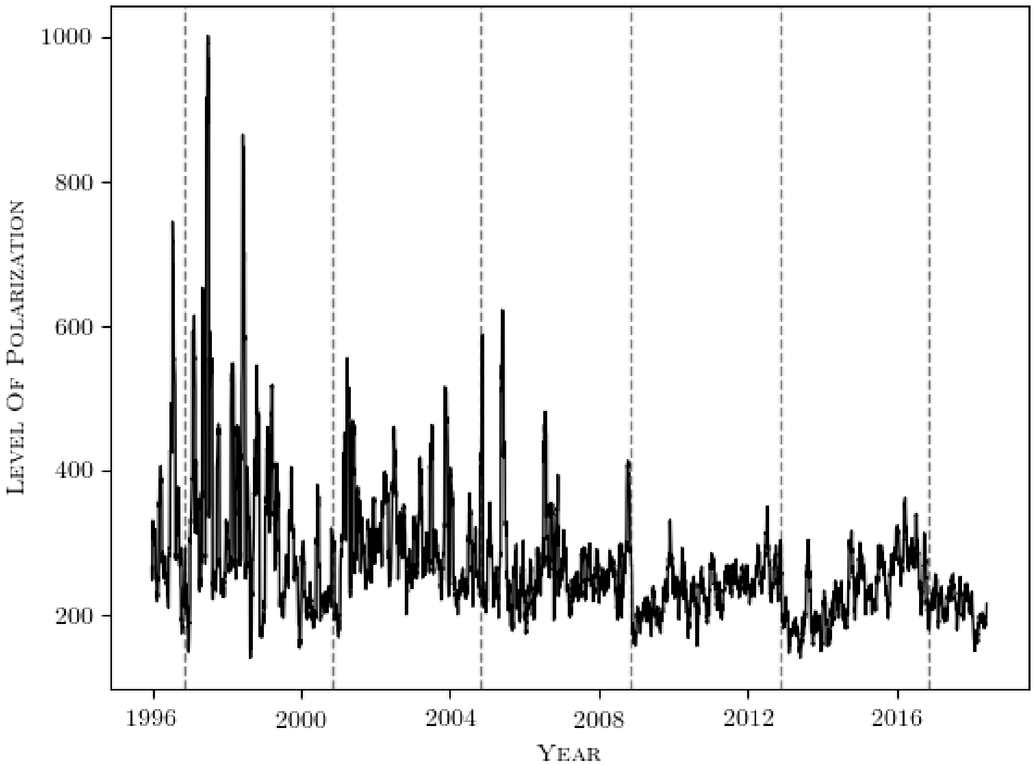
4.1 Descriptive statistics

A visual representation of the trajectory of the polarization metric is shown in Figure 4 below, while the corresponding values can be found in Appendix C. Looking at both these describing displays of the polarization scores, there seems to be a noisy downward trend in the level of polarization till 2009. Especially in the first years of the studied timespan, the volatility in the levels of polarization is high. This corresponds with the standard errors in the article scores as can be seen in the previous section.

The grey striped lines show the presidential election days from 1996 till 2016. In the latter three elections, there seems to be a pattern of a steep decrease starting just before the election day. This could be due to the fact that the lead days already start to influence the polarization score just before election day. Further, this cycle shows a long period of a steady rise after this steep decline. The elections of 1996, 2000 and 2004 do not show this pattern but are rather dominated by a lot of noise in the trend.

Figure 4. Trend of the daily polarization metric over time

The level of polarization is calculated using a moving variance with the windows -14 to +14 on the day of interest on the ideological scores of the articles. Note that the base levels of polarization do not bear substantial meaning, but that the trend rather indicates the movement in the level of polarization in the public. The grey vertical striped lines show the presidential election days in the United States.



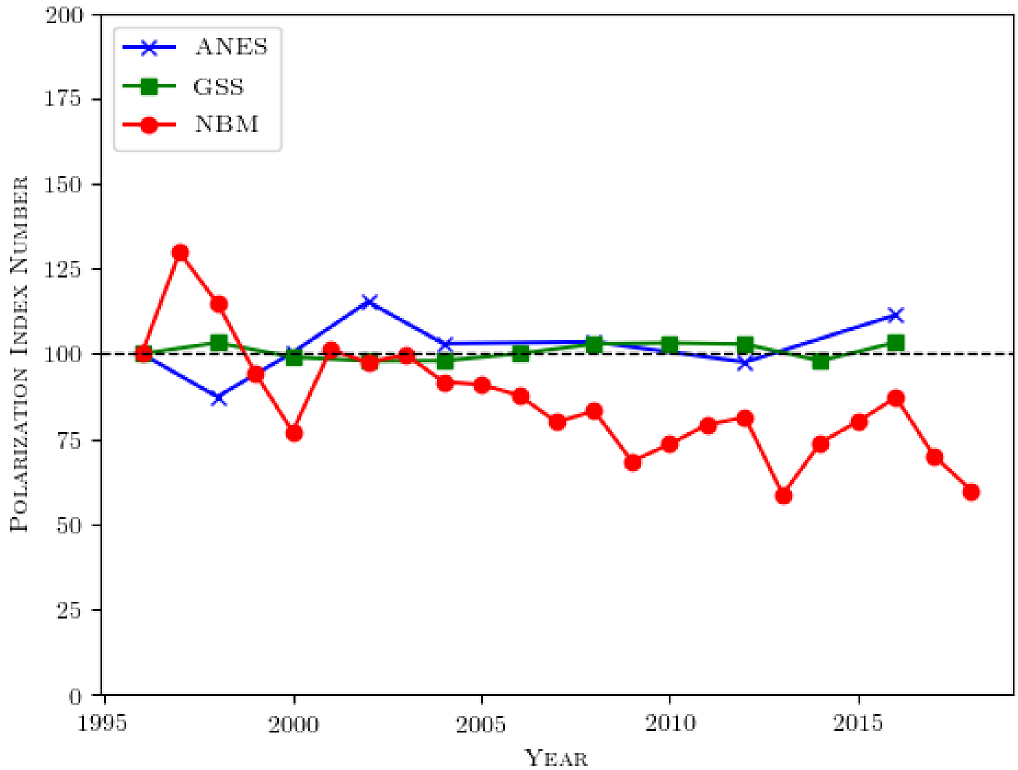
4.2 Empirical results

Validation of the metric

A graphical representation of the trajectory of the different polarization metrics as described in Section 3.4 are shown in Figure 5 below. This indicates that both the ANES and GSS polarization metric, the bimodality coefficient of the survey scores, remain stable around the indexed number for the 1996 polarization. These values are similar to the bimodality coefficients calculated by Lelkes (2016) on the same data. Contrarily, the new metric developed in this study seems to show a decreasing pattern. Even though they are hypothesized to measure the same principle.

Figure 5. Comparison of different polarization measurements

This figure contains the different polarization metrics over time. Respectively polarization metric calculated using the American National Election Survey (ANES) data, General Social Survey (GSS) data, and the polarization metric as introduced by current paper, called the Newspaper Based Metric (NBM) in this regard. Note that the ANES and the GSS do not have yearly data, the collection frequency by these institutions is lower. The different metrics are all set to 100 at their observation in 1996 and changes in the metric are processed by calculated based on their previous index number.



The results of the correlations displayed in Table 3 confirm our visual conclusion that the trend of the newly formed metric is not captured by the trends of both survey-based metrics. The pairwise correlation coefficient is insignificant in relation to both the ANES- as well as the GSS data. Surprisingly, also the coefficient between the ANES- and the GSS-data is insignificant. However, as conceptually the developed metric is so different from the survey-based methods it is plausible that the comparison between the newspaper-based metric and the survey-based method is not informative. Therefore, several justifications can be made to explain why both these different techniques to look at the same principle could exhibit useful results.

Table 3. Correlations between different polarization metrics

Correlation coefficients to denote the relationships between the polarization metrics originated from the ANES data, the GSS data and the developed way of calculating the polarization score, the news-based metric. Standard errors are denoted in parentheses.

	GSS data	ANES data	News-Based data
GSS data	1		
ANES data	-0.3778 (0.3561)	1	
News-Based data	0.616 (0.8572)	-0.3607 (0.3800)	1

*, ** and *** represent the statistical significance at the 10%, 5% and 1%, respectively.

First, the frequency of the survey data is low. This makes formal comparison unreliable and also gives the visual illusion of different trends while effectively it is based on solely 11 (GSS) or 8 (ANES) observations. The low frequency also presents a matching issue as the GSS is conducted from mid-April to the end of August while the ANES is conducted from early-September to January. However, news articles are used from all year around which might arouse comparability problems. As events may be newsworthy in March, they may still influence the thoughts of people in May. Hence, excluding such events from the metric might rise different concerns. Especially as the boundaries of such a limited window are vague and hence arbitrary. Second, it is not clear that measuring public polarization using survey data yields the optimal results. As scientific research is not yet unanimous on how to measure polarization, the results of any correlational evidence would be dependent on which polarization metric is used as a reference. It might be the case that when using one of the four polarization metrics as developed by Lelkes (2016), a significant correlation is found. Also, the main advantage of current metric is the fact it yields daily results. The daily change cannot be tested as there is no metric that has data at the same frequency. This shows that conceptually, the developed metric is so different in measuring polarization from the previously used survey-based methods that, by definition, comparison is unlikely to be informative.

In conclusion, both formal and visual evidence shows that the developed metric does not follow previously used polarization metrics. However, based on the fact that the frequency of the test data is low, there is no perfect metric and that the newspaper metric yields distinct advantages, the newspaper-based methodology might capture different aspects of polarization. Especially since the ANES and GSS data are unlikely to capture the temporal dynamics in polarization which the newspaper-based method may do. Ultimately, further research into validating the metric is necessary but using it might still provide useful insights into polarization in the American public. Hence, we do not accept nor reject our constructed H_1.

Trends in the level of polarization

As determined in Section 4.1, visual evidence seems to show a decreasing pattern up to 2009 and afterwards a cyclical pattern till the end of the timespan. Hence, to formally assess whether this is the case, Newey-West regressions are run on the full timespan as well as on the partial timespan ranging from January 2009 till May 2018.

The results of the regression analysis can be found in Table 4. Regression (1) and (2) show the results on the total timespan. As visual interpretation of Figure 4 already hinted, the polarization trend over time is significantly downward sloping when looking at the complete sample. In both the simple univariate regression and the multivariate extended regression, the coefficient for time is negative and significantly different from 0 at the 1 % level. This indicates that there is strong evidence that the level of polarization is decreasing based on the discourse used in *The New York Times*. Strikingly, all coefficients for the categorical variable "*Executive Editor*" are highly significant and negative. This would mean that Dean Baquet has had a statistically significant positive effect on the polarization levels being the reference category. As Dean Baquet took over the position of executive editor around 2014, the trend is somewhat set to a higher point on the trendline. The most plausible explanation for this result would be to that this variable is an offset for the decreasing factor of time over the years. Interestingly, there seems to be a monthly seasonality effect in the level of polarization. June and July see higher polarization levels and are significant at the 5% and 1% level. This could be due to Congressional recess in August, as just before recess, politicians try to pass bills causing more ideological clashes. Also, the significantly negative coefficient for December could reflect common America emphases on family, Santa Claus and charity to the poor in this month (Golby and Purdue, 1986). There might also be an offset of the presidential elections being in November once every four years. A strong decrease in polarization could reflect the sudden drop in political debates. The coefficient for the return on the NYT stock is insignificant while having a Republican president significantly (at the 5% level) decreases the polarization metric.

As visual evidence shows that the clear downward trend seems to be ending around 2009, performed are two more regressions. Regression (3) and (4) in table 4 are restricted to observations from January 1, 2009 to May 31, 2018. In both regressions, the polarization trend over time is statistically insignificant. This shows that the decreasing pattern halted, and the polarization metric stabilizes around the year 2009. Notably in equation (4), both the effect of the (remaining) changes in executive director and party affiliation of the US president remain negative in direction and statistically significant.

Concluding, the evidence shows a clear downward sloping trend in the level of polarization in the US society in the total timeframe in this study. More specifically, while controlling for external influences, the metric decreased over time until the year 2009 while afterwards it stabilized.

Table 4. Polarization Over Time

This table contains the results of regressing the polarization metric on time. The dependent variable is the absolute polarization metric. Regression (1) and (2) show the regressions on the whole timespan, both condensed as well as extended. Regression (3) and (4) show the results of regressions from 2009 till 2018. The variable of interest is the *date* variable, to determine whether there exists a trend over time. *Executive Editor* is referenced to the current executive editor, Dean Baquet. *Month* is referenced to January, party affiliation president is a dummy with 0 = democrat and 1 = republican. Standard errors are denoted in parentheses and are robust to heteroskedasticity.

	1996 - 2018		2009 - 2018	
	(1)	(2)	(3)	(4)
Date	-0.017***	-0.0331***	0.002	-0.008
	(0.002)	(0.005)	(0.002)	(0.007)
Executive Editor				
Jill Abramson		-59.663***		
		(10.135)		
Bill Keller		-81.921***		
		(15.538)		
Joseph Lelyveld		-123.040***		-39.565**
		(31.637)		(15.504)
Howell Raines		-84.983***		-37.900***
		(25.382)		(9.910)
Return on "NYT Company"		-35.556		-27.534
		(33.780)		(23.215)
Party Affiliation President		-14.490**		-35.229***
		(7.167)		(7.169)
Month				
February		3.399		-23.886***
		(10.572)		(7.246)
March		12.873		-10.960
		(10.135)		(9.154)
April		-2.013		-13.116*
		(9.564)		(7.845)
May		15.109		-19.988**
		(12.590)		(8.205)
June		49.742**		-7.236
		(19.543)		(9.376)
July		47.025***		5.339
		(14.373)		(10.285)
August		-0.289		-14.488
		(9.686)		(9.654)
September		-1.169		-4.753
		(8.112)		(8.187)
October		8.388		-6.881
		(10.027)		(7.382)
November		3.651		-4.863
		(11.130)		(9.664)

December		-26.948***		-8.865
		(9.602)		(23.215)
Constant	565.250***	914.616***	199.699***	421.271***
	(32.383)	(107.552)	(43.203)	(147.983)
Observations	8,187	8,186	3,438	3,438
Adjusted R^2	0.1888	0.2680	0.0013	0.2010

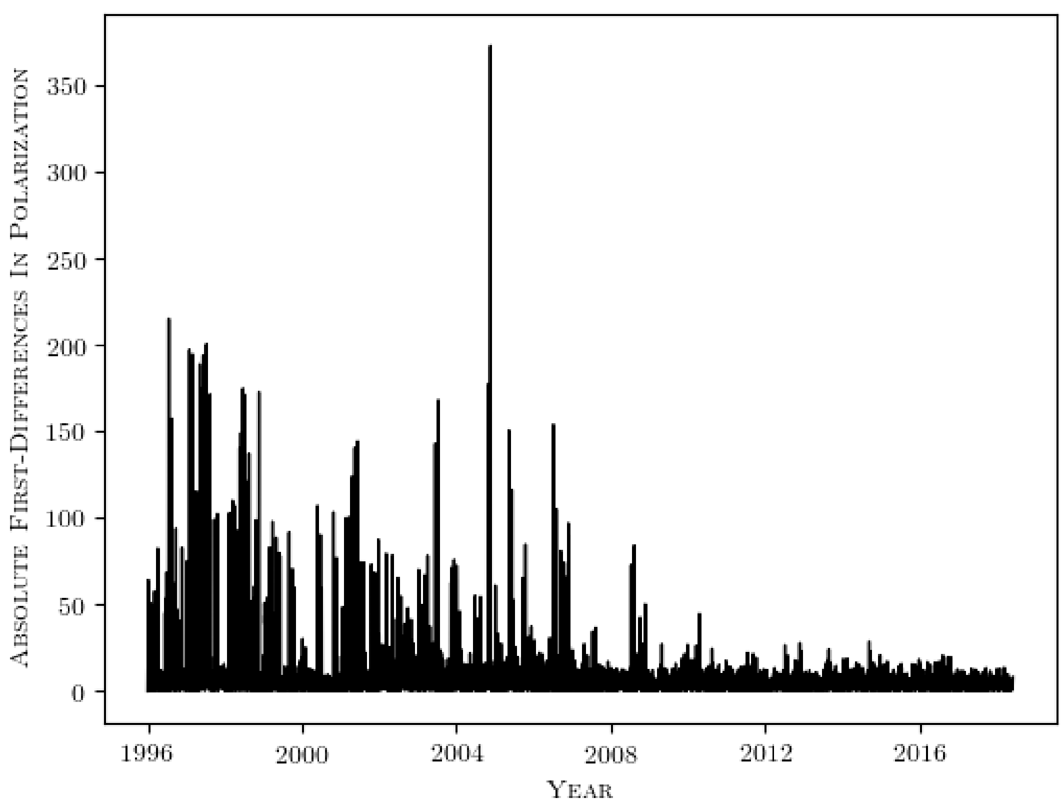
*, ** and *** represent the statistical significance at the 10%, 5% and 1%, respectively.

Volatility of the polarization metric

As designated in the descriptive statistics, the spread between observations in the polarization metric seems to decrease. This is substantiated by looking at the Figure 6 below. The plotted first differences over time show increasingly less spikes as the time goes on. Especially in the first years of the sample, it appears that the measurements are able to change drastically in small periods of time. To formally assess whether this development is significantly present in the data, Newey-West regressions on the absolute first-differences on time are both run on the full sample as well as on the partial timespan from 2009 till 2018.

Figure 6. Absolute first differences plotted against time

This first differences are the daily changes between subsequent daily polarization levels. The differences are made positive by taking the absolute value of the differences.



The results of these regressions can be found in Table 5. These results show a significant decrease in the absolute first-differences of the developed polarization metric over time. In the full timespan, both the condensed as well as the extended equation show negative coefficients which are significant at the 1% level. In the partial timespan, the insignificant coefficient in the extended equation indicates that the spread in observations is no longer decreasing. Within the control variables, negative and significant coefficients are obtained for all the executive directors compared to the current executive director. However, looking at the visual representation in Figure 1, it seems unlikely that the appointment of Dean Baquet – on May 14, 2014 – has spiked a new volatile period in the article scores. Looking at the evidence of the partial timespan, it is the case that the first-differences do not decrease anymore after 2009. Hence the positive effect of the appointment of Dean Baquet on the polarization metric could solely be offsetting the constant decrease caused by the decreasing coefficient on the date variable. This is supported by the insignificant coefficients in the partial timespan on the coefficients for the two remaining categories for executive editor. Further, the remainder of the control variables do not show consistent significant values

So, there is strong evidence of decreasing first differences and hence a decreasing trend in the day to day changes in the polarization metric. The decrease is significant at the 1% level, however, when solely the latter timespan is studied, the decreasing effect is absent. This indicates that the volatility decreases and afterwards stabilizes.

Table 5. Absolute changes in polarization regressed over time

This table contains the results of regressing the absolute first differences of the polarization metric on time. Regressions (1) and (2) show the results on the total timespan, from January 1996 to May 2018. Regressions (3) and (4) show the results of regressing on the partial timespan, from January 2009 to May 2018. Added are the identical controls as used in Table 4. Standard errors are denoted in parentheses and are robust to heteroskedasticity.

	1996 - 2018		2009 - 2018	
	(1)	(2)	(3)	(4)
$Date_t$	-0.001***	-0.002***	-0.0002***	0.000
	(0.0001)	(0.0004)	(0.0001)	(0.0002)
Executive Editor				
Jill Abramson		-2.139***		
		(0.582)		
Bill Keller		-3.258***		
		(0.980)		
Joseph Lelyveld		-4.461**		-2.400
		(2.097)		(0.287)
Howell Raines		-4.625***		0.196
		(1.668)		(0.491)
Return on "NYT Company"		0.158		-0.706
		(5.959)		(2.915)

Party Affiliation President		-0.653		-0.860***
		(0.474)		(0.256)
Month				
February		0.265		-0.087
		(0.639)		(0.309)
March		0.831		0.108
		(0.753)		(0.307)
April		0.237		0.415
		(0.656)		(0.335)
May		0.931		-0.178
		(0.789)		(0.270)
June		2.565**		0.068
		(1.241)		(0.249)
July		2.752**		0.458
		(1.282)		(0.320)
August		0.801		0.343
		(0.696)		(0.332)
September		-0.572		0.248
		(0.575)		(0.327)
October		0.549		0.203
		(0.655)		(0.272)
November		0.957		0.034
		(0.972)		(0.370)
December		-0.697		0.183
		(0.595)		(0.280)
Constant	24.867***	38.074***	8.301***	3.772
	(2.134)	(7.342)	(1.463)	(4.873)
Observations	8,186	8,186	3,438	3,439
Adjusted R^2	0.029	0.0368	0.0036	0.0060

*, ** and *** represent the statistical significance at the 10%, 5% and 1%, respectively.

Application: The effect of the election cycle

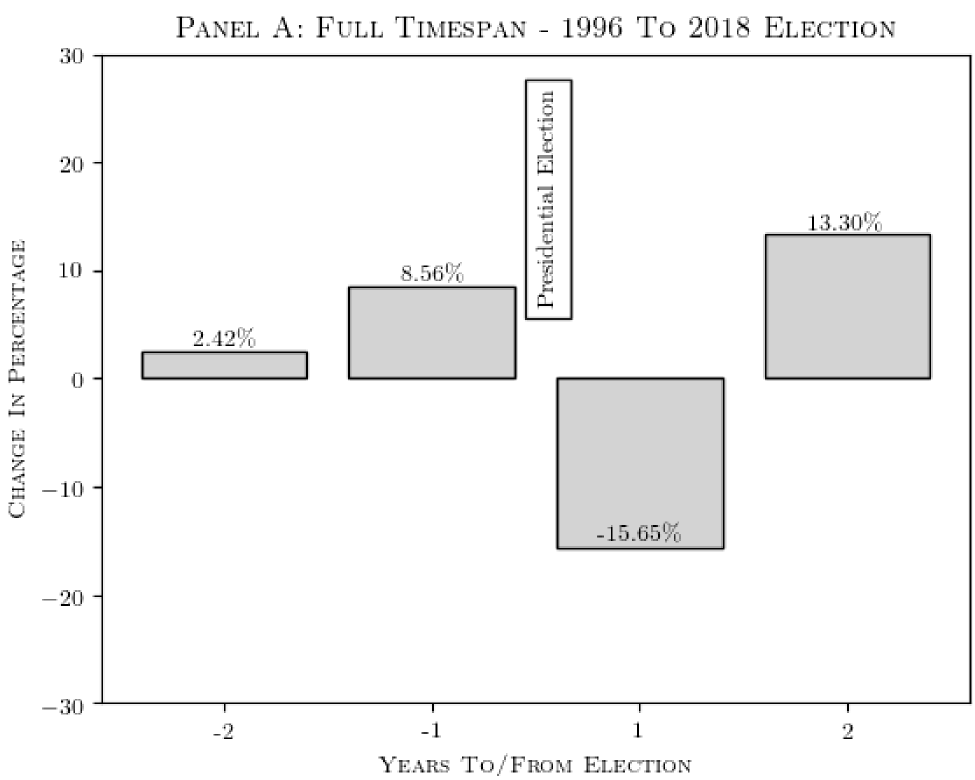
Figure 7 shows the change per year in the election cycle as described in Section 3.4.4. Visible in Panel A – encompassing all election cycles in the sample – is that on average, the year after the presidential elections, the polarization metric decreases with a total of 15.65 %. Followed by rises of on average 13.30 %, 2.42% and 8.56% up to the presidential elections. A similar result is found for the last 3 elections as displayed in Panel B. Panel B shows a similar trend when comparing it to Panel A: 3 years of rising polarization levels up to the election followed by a sharp decline in polarization levels the first year after the presidential elections.

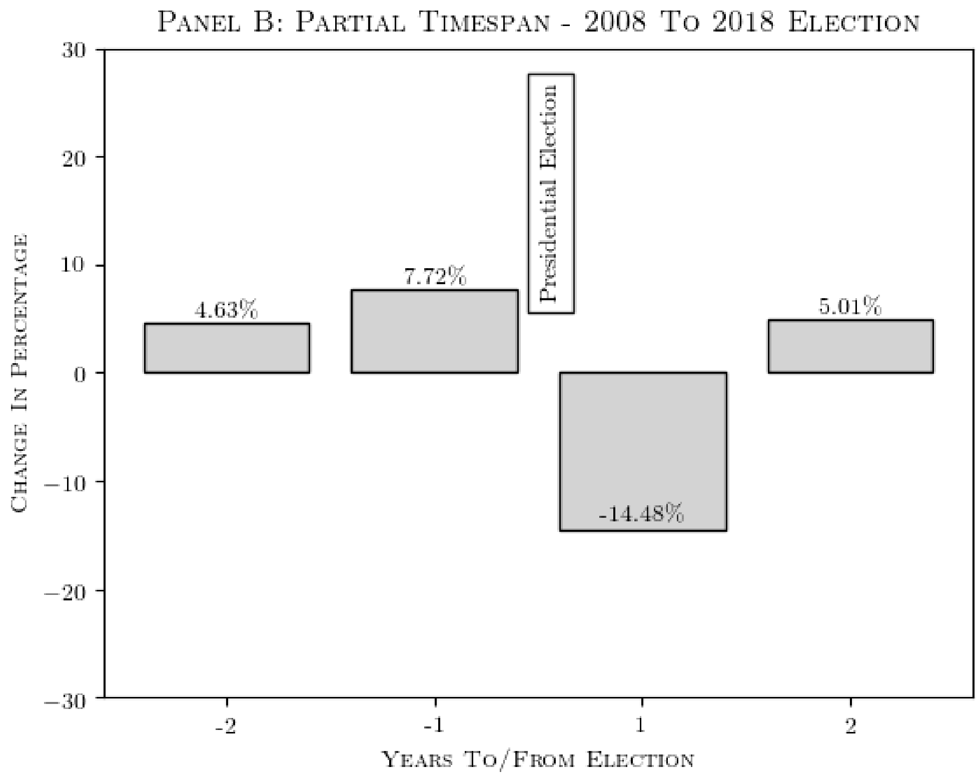
A Mann-Whitney U test is performed to formally determine whether – as the interpretation of the visual evidence suggests – the election year has a significant influence on the change in polarization. The sample is divided into two groups: (a) the change in polarization one year after the election (b) all other years. On the complete sample, The Mann-Whitney U test shows that there is a statistically significant difference in the development in polarization between the four groups, z = 1.750, p = 0.0801. As the p-value < 0.10, the effect is statistically significant at the 10% level. On the partial sample with observations from the 2008 to the 2016 elections, the Mann-Whitney U test does not show a statistically significant differences between the election years, z = 1.387, p = 0.1655. So, the differences in changes in polarization are not sufficient to be statistically significant.

So, there is graphical evidence that in the years building up to a presidential election the polarization rises and in the year after the election it decreases again. Formal testing shows that there is minor evidence of a significant difference in the distributions between the years in the election cycle. Using Mann-Whitney U tests shows that in the full sample, there is a significant difference (at the 10% level) between the distributions. Running the test on the partial timespan – from 2009 till 2018 – does not return significant and hence the distributions is are not different.

Figure 7. Changes in Polarization over Presidential Election Cycle

Panel A shows the changes in polarization over the presidential election cycle over our total sample. Panel B shows the changes in polarization over the presidential election from the 2008 election to the 2018 election. Note that available data prohibits current study from (completely) including the 1996 and 2018 elections[12]. Percentages are calculated as the average of the changes from the first day to the last in the respective election cycle year.





5 Discussion and conclusion

The aim of this study was to develop a new metric to measure the level of polarization in the American public. As increasingly frequent used to describe as a cause of societal issues by policy makers, politicians and newspapers, there is little guidance in what polarization actually means and how it should be measured. The developed metric analyzes the discourse in political news articles published in *The New York Times* from January 1996 to May 2018 by applying the *Wordscores* algorithm. Using the individual articles scores to calculate a moving variance for each day, a daily polarization mark is obtained.

Following previous studies that newspaper data may be used to proxy the views of the public, the outcomes of the metric are compared to measurements of polarization. The results reveal that there is no significant correlation between the newspaper-based metric with the metric making use of GSS and ANES data. As it is the cornerstone of this study, it raises questions whether the results of this research actually measure what they intend to. However, also the estimates between the ANES and GSS – both making use of survey data on the same target audience – have no statistically significant correlation in its estimates. That demonstrates the complexity in quantifying this comprehensive concept and presents the need for more reliable techniques. Simultaneously, it shows the fact that even though the estimates are not correlated with other measurements of the subject, they still might bear meaning in capturing a different aspect of the broad concept “polarization”. Especially since the developed metric yields distinct advantages compared to the survey-based determination of polarization. Primarily, relevant for scholars, having a daily metric gives ample opportunities to study the causes and effects of changes in the level of polarization. Also, more practically, it enables institutions and governmental bodies to review the impact of implemented regulations on a short notice as well as whether the effect of events such as sport successes (Van Hilvoorde et al., 2010) or terrorist attacks (Berrebi and Klor, 2008) influence the polarization level in a society. Such events could have implications for polarization based on changes in national cohesion and group feel. On a more practical note, after initial setting up the computations, the metric is both substantially more cost-effective as well as time-effective compared to survey-based methods what increments its usability.

The evidence found on the trend in the levels polarization in the United States is rather surprising. Both graphical evidence as well as regression analyses shows a decreasing level in polarization from 1996 till 2009 based on the evidence collected from *The New York Times*. Afterwards, the trend seems to stabilize, and a cyclical pattern emerges. The cyclical pattern is characterized by multiple years of slow rise followed by a steep decline. These results confirm the hypothesis that the level of polarization does not rise in the American public. However, the evidence towards a decline is unexpected as it exceeds the conclusions of skeptics of the rising-polarization belief. Further, noticeable is the strong evidence towards the decreasing effect on polarization if a republican president holds the Oval Office. This might be due to the ideological asymmetry in the American party system. Republicans are more unified, even on issues they do not all agree upon what smoothens both the implementation of policies as well the news regarding it if they are in office (Lelkes and Sniderman, 2016). More detailed analysis into these assumptions might yield interesting results for political scholars.

Also, the results found on the volatility of the metric are consistent with the evidence that newspapers are more susceptible to the wishes of its audience, and at the same time are better at tracking those wishes. That is, the volatility of the polarization metric decreases statistically significant until the 2009 mark. Afterwards, it remains at similar levels to the end of the sample, until May 2018. This indicates that the metric increases in stability as well as reliability as it is unlikely that the aggregated polarization levels in the US have a lot of sudden shifts. Also, it could demonstrate that the trend as before 2009 is not representative of the US population due to the inability of newspapers to follow the wishes of the audience. It would be interesting to see if the high-volatile period keeps increasing before 1996, as that would substantiate this claim. In regard to current methodology, developments in the media sector might negate the usefulness of this metric in the near future. Decreasing sales has already bankrupted certain newspapers, and with the continuous growth of internet sources, the sales levels might drop to a level that the ideological views of the public are not reflected by newspaper discourse anymore.

To showcase a potential application of the high-frequency metric the presidential election cycle effect – primarily known from stock-market returns – is studied. Based on theory the level of polarization should rise up the presidential election, after which it should decrease. Formal tests to measure the changes in polarization in regard to the amount of years up to the closest presidential election show that this effect is significant in the complete sample period. That is, in the elections from 1996 to 2016, the first year after an election shows a significant different yearly change compared to the other years in the cycle. This effect is absent in the partial sample from the 2008 elections to the 2016 elections. This could be to the fact that the amount of cycles is relatively small and the differences in changes have to be substantial to be significant. Since the cyclical pattern seems to strengthen after 2009, more data might be needed to signify the effect. Otherwise, there might be a different phenomenon responsible for the cyclicity in the polarization metric which would need further examination.

As discussed, primary focus for future research should be on improving and validating the polarization metric itself. Regardless of feasibility, using the articles of all newspapers in the US weighted on audience ratings would be optimal to proxy the views of the American public. Further, as machine learning gains more recognition as a primary way to analyze texts, other algorithms might improve the extracting of the ideological position out of texts by the *Wordscores* algorithm. Even though not yet applicable to large quantities of articles, the *Wordfish* algorithm by Slapin and Proksch (2008) shows promising results as it is not dependable of the availability of suitable reference texts. Hence, it completely mitigates any potential biases in human decision-making. Also, it might be interesting to compare multiple ways to measure polarization out of ideological positions. Comparing approaches such as bimodality, variance and overlap coefficients might result in a universal, optimal way to arrive at a polarization level out of ideological positions. As it is currently absent, it will increase the comparability of studies and provide guidance to future scholars. Once validated and generally agreed upon, the opportunities for research a daily polarization metric yields are plentiful. Analyzing it in respect to other social-economic factors such as immigration rates, the passing of new bills in the senate or presidency approval rates might yield new insights in the movement of polarization over time. Finally, this might show opportunities for policy makers to positively impact the level of polarization.

Encompassing the essence of current research, this study has developed a new way to measure polarization. Even though it still needs further developing, it might provide a useful genesis in generating a daily specification of polarization.

Bibliography

Abramowitz, A. 2010. *The disappearing center: Engaged citizens, polarization and American democracy*. New Haven, CT:: Yale University Press.

Abramowitz, A.I. and Saunders, K.L. 2008. "Is polarization a myth?" *The Journal of Politics 70(2)* 542-555.

Allvine, F.C. and O'Neill, D.E. 1980. "Stock market returns and the presidential election cycle: Implications for market efficiency." *Financial Analysts Journal* 49-56.

Alwin, D.F. and Tufiş, P.A. 2016. "The changing dynamics of class and culture in American politics: A test of the polarization hypothesis." *The Annals of the American Academy of Political and Social Science 663(1)* 229-269.

Andrews, K.T. and Caren, N. 2010. "Making the news: Movement organizations, media attention, and the public agenda." *American Sociological Review 75(6)* 841-866.

Antilla, L. 2005. "Climate of scepticism: US newspaper coverage of the science of climate change." *Global Environmental Change 15(4)* 338-352.

Baek, K., Holton, A., Harp, D. and Yaschur, C. 2011. "The links that bind: Uncovering novel motivations for linkign on Facebook." *Computers in Human Behavior 27(6)* 2243-2248.

Baldassarri, D. and Bearman, P. 2007. "Dynamics of political polarization." *American Sociological Review 75(2)* 784-811.

Barranco, J. and Wisler, D. 1999. "Validity and systematicity of newspaper data in event analysis." *European Sociological Review 15(1)* 301-322.

Baumgarten, B. and Grauel, J. 2009. "The theoretical potential of website and newspaper data for analysing political communication processes." *Historical Social Research 34(1)* 94-121.

Baumgartner, J.C. and Morris, J.S. 2010. "MyFaceTube politics: Social networking web sites and political engagement of young adults." *Social Science Computer Review 28(1)* 24-44.

Baur, N. and Lahusen, C. 2005. "Sampling process-generated data in order to trace social change." *Recent Developments and Applications in Social Research Methodology*.

Bennett, L. 1988. *News: The Politics of Illusion*. New York and London: Longman.

Benoit, K. and Laver, M. 2008. "Compared to what? A comment on "A robust transformation procedure for interpreting political text" by Martin and Vanberg." *Political Analysis 16(1)* 101-11.

Berrebi, C. and Klor, E.F. 2008. "Are voters sensitive to terrorism? Direct evidence from the Israeli electorate." *The American Political Science Review 102(3)* 279-301.

Budge, I. and Pennings, P. 2006. "Do they work? Validating computerised word frequency estimates against policy series." *Electoral Studies 26(1)* 121-129.

Corbett, J.B. and Durfee, J.L. 2004. "Testing public (un)certainty of science: Media representations of global warming." *Science Communication 26(2)* 129-151.

DiMaggio, P., Evans, J. and Bryson, B. 1996. “Have American's social attitudes become more polarized?” *American journal of Sociology 102(3)* 690-755.

Dimmick, J., Chen, Y. and Li, Z. 2004. “Competition between the Internet and Traditional News Media: The gratification-opportunities niche dimension.” *The Journal of Media Economics 17(1)* 19-33.

Earl, J., Martin, A., McCarthy, J.D. and Soule, S.A. 2004. “The use of newspaper data in the study of collective action.” *Annual Review of Sociology 30(1)* 65-80.

Entman, R.M. 1989. “How the media affect what people think: An information processing approach.” *The Journal of Politics 51(2)* 347-370.

Evans, J.W. 2003. “Have Americans' attitudes become more polarized? - An update.” *Social Science Quarterly 84(1)* 71-90.

Fiorina, M.P and Abrams, S.J. 2008. “Political polarization in the American public.” *Annual Review of Political Science* 563-588.

Fiorina, M.P, Abrams, S.J. and Pope, J. 2005. *Culture war? The myth of a polarized America.* New York: Pearson Longman.

Freeman, J.B. and Dale, R. 2013. “Assessing bimodality to detect the presence of a dual cognitive process.” *Behavior Research Methods 45(1)* 83-97.

Galambos, L. 1975. *Public image about big business, 1880-1940: A quantitative study in social change.* Baltimore, MD: John Hopkins University Press.

Gans, H.J. 1979. “The messages behind the news.” *Columbia Journalism Review 17(5)* 40-45.

Golby, J.M. and Purdue, A.W. 1986. *The Making of the Modern Christmas.* Athens, GA: University of Georgia Press.

Hansen, L.P. and Hodrick, R.J. 1980. “Forward exchange rates as optimal predictors of future spot rates: An econometric analysis.” *Journal of Political Economy 88(5)* 829-853.

Hetherington, M.J. 2001. “Resurgent mass partisanship: The role of elite polarization.” *American Political Science Review 95(3)* 619-631.

Hetherington, M.J. 2008. “Turned off or turned on? How polarization affects political engagement.” *Red and Blue Nation? Volume 1: Characteristics and Causes of America's Polarized Politics.* 1-33.

Hyndman, R.J. and Athanasopoulos, G. 2018. *Forecasting: Principles and practice.* OTexts.

Iyengar, S., Sood, G. and Lelkes, Y. 2012. “Affect, not ideology: A social identity perspective on polarization.” *Public Opinon Quarterly 76(3)* 405-431.

Jacobsen, G.C. 2000. “Party polarization in national politics: The electoral connection.” In *In Polarized Politics: Congress and the President in a Partisan Era*, by J.R. and Fleisher, R. Bond, 9-30. Washington, DC: CQ Press.

Johnson, T.J. and Perlmutter, D.D. 2010. “Introduction: The Facebook election.” *Mass Communication and Society 13(5)* 554-559.

Laver, M., Benoit, K. and Garry, J. 2003. “Extracting policy positions from political texts using words as data.” *American Political Science Review 97(2)* 311-332.

Layman, G.C. and Carsey, T.M. 2002. “Party polarization and "conflift extension" in the American electorate.” *American Journal of Political Science 46* 786-802.

Lelkes, Y. and Sniderman, P.M. 2016. “The ideological asymmetry of the American party system.” *British Journal of Political Science 46(4)* 825-844.

Lelkes, Y. 2016. “Mass polarization: Manifestations and measurements.” *Public Opinon Quarterly 80(1)* 392-410.

Levendusky, M. 2009. *The partisan sort: How Liberals became Democrats and Conservatives became Republicans.* Chicago, IL: University of Chicago Press.

Mann, H.B. and Whitney, D.R. 1947. “On a test of whether one of two random variables is stochastically larger than the other.” *Annals of Mathematical Statistics 18(1)* 50-60.

Martin, L.W. and Vanberg, G. 2008. “A robust transformation procedure for interpreting political text.” *Political Analysis 16(1)* 93-100.

McCarty, N., Poole, K. and Rosenthal, H. 2006. *Polarized America.* Cambridge, MA: MIT Press.

McCombs, M. 2002. “The agenda-setting role of the mass media in the shaping of public opinion.” *Paper presented at the London School of Economics Conference on Mass Media, London.*

Meraz, S. 2011. “The fight for 'how to think': Traditional media, social networks, and issue interpretation.” *Journalism 12(1)* 107-127.

Mouw, T. and Sobel, M.E. 2001. “Culture wars and opinion polarization: the case of abortion.” *American Journal of Sociology 106(4)* 913-943.

Myers, D.J. 2000. “The diffusion of collective violence: Infectiousness, susceptibility and mass media networks.” *Americcan Journal of Sociology 106(1)* 173-208.

Napoli, P.M. 2010. *Audience evolution: New technologies and the transformation of media audiences.* New York: Columbia University Press.

Newey, W.K. and West, K.D. 1987. “A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix.” *Econometrica 55(3)* 703-708.

Oliver, P.E. and Myers, D.J. 1999. “How events enter the public sphere: Conflict, location and sponsorship in local newspaper coverage of public events.” *American Journal of Sociology 105(1)* 38-87.

Ortiz, D.G., Myers, D.J.,Walls, E. and Diaz, M.D. 2005. “Where do we stand with newspaper data?” *Mobilization 10(3)* 397-419.

Singer, M.M. and Carlin, R.E. 2013. “Context counts: The election cycle, development, and the nature of economic voting.” *The Journal of Politics 75(3)* 730-742.

Slapin, J.B. and Proksch, S. 2008. “A scaling model for estimating time-series party positions from texts.” *American Journal of Political Science 52(3)* 705-722.

Theriault, S. 2008. *Party polarization in the congress.* New York: Cambridge University Press.

Van Hilvoorde, I., Elling, A. and Stokvis, R. 2010. "How to influence national pride? The Olympic medal index as unifying narrative." *International Review for the Sociology of Sport 45(1)* 87-102.

Vitak, J., Zube, P., Smock, A., Carr, C.T., Ellison, N. and Lampe, C. 2011. "It's complicated: Facebook users' political participation in the 2008 election." *CyberPsychology, behavior, and social networking 14(3)* 107-114.

Vu, H.T. 2014. "The online audience as gatekeeper: The influence of reader metrics on news editorial selection." *Journalism 15(8)* 1094-1110.

Weaver, D.H. 2007. "Thoughts on agenda setting, framing, and priming." *Journal of Communication 57(1)* 142-147.

Weingart, P., Engels, A. and Pansegrau, P. 2000. "Risks of communication: Discourses on climate change in science, politics, and the mass media." *Public Understanding of Science 9(3)* 261-284.

Wong, W. and McAleer, M. 2009. "Mapping the presidential election cycle in US stock markets." *Mathematics and Computers in Simulation 79(11)* 3267-3277.

Woolley, J.T. 2000. "Using media-basede data in studies of politics." *American Journal of Political Science 44(1)* 156-173.

Zhao, X., Liano, K. and Hardin III, W.G. 2004. "Presidential election cycles and the turn-of-the-month effect." *Social Science Quarterly 85(4)* 958-973.

Appendixes

Appendix A. Comparative Manifesto Project (CMP) ideological reference scores used for calculating the scores for each individual word. Lower numbers signal a more liberal view while higher numbers signal a more conservative view.

Election Year	Democratic Party Platform	Republican Party Platform
1996	8.784	24.217
2000	-3.596	33.314
2004	8.361	25.903
2008	11.14	25.124
2012	-6.442	27.957
2016	-20.578	32.969

Appendix B. Descriptive statistics of the article scores split on publication year. The data is gathered by applying the *Wordscores* algorithm on the political news articles in *The New York Times*. Note that individual scores do not bear meaning, rather does the trend in scores.

	Obs.	Mean	Std. Dev.	Min.	Median	Max.
Total Sample	78,314	11.28	17.29	-156.40	11.26	88.79
1996	3,783	14.11	17.94	-125.68	13.49	69.97
1997	2,598	16.70	20.43	-144.05	17.35	65.13
1998	3,085	13.25	19.23	-150.51	13.03	66.31
1999	2,787	15.05	17.44	-117.52	14.65	67.35
2000	4,589	12.96	15.75	-142.72	12.43	67.51
2001	3,069	17.16	18.04	-139.10	17.90	71.65
2002	2,872	16.24	17.71	-126.70	16.20	81.92
2003	2,848	16.53	17.91	-119.55	17.32	79.18
2004	3,901	12.15	17.18	-156.40	11.85	73.81
2005	2,517	17.49	17.10	-118.84	18.72	87.79
2006	2,564	14.96	16.81	-136.60	15.47	88.79
2007	1,943	12.51	16.04	-60.41	12.55	67.62
2008	3,874	5.81	16.36	-130.85	5.24	67.42
2009	2,706	15.56	13.83	-48.43	16.02	70.57
2010	2,917	11.78	15.37	-60.38	11.63	64.02
2011	4,719	10.53	15.96	-71.22	10.56	65.31
2012	5,806	5.89	16.17	-78.73	5.15	64.01
2013	2,916	15.85	13.73	-40.44	16.31	59.60
2014	3,082	11.20	15.42	-56.06	11.42	65.60
2015	4,531	6.75	16.06	-70.02	6.12	66.97
2016	4,857	-0.21	16.75	-61.13	-1.77	60.82
2017	3,559	7.82	15.02	-41.20	7.08	58.08
2018	1,791	5.62	13.87	-50.68	4.92	49.06

Appendix C. Descriptive statistics of the polarization scores split on both the full sample as well as on a yearly basis. The values are obtained by calculating a moving variance of -14 days to +14 days on the obtained article scores.

	Obs.	Mean	Std. Dev.	Min.	Median	Max.
Total Sample	8,187	271.32	92.71	139.57	248.47	1000.48
1996	365	305.99	112.38	148.20	272.55	743.25
1997	365	414.42	197.15	187.76	349.80	1000.48
1998	365	367.33	157.62	139.87	332.16	863.89
1999	365	299.27	80.69	154.00	279.51	517.23
2000	365	235.80	45.60	177.54	220.64	379.01
2001	365	325.86	83.26	168.65	304.41	555.02
2002	365	312.39	54.38	200.33	304.88	459.81
2003	365	315.29	72.24	220.46	289.79	512.84
2004	365	276.42	70.40	200.48	244.33	586.81
2005	365	282.45	98.76	177.47	253.36	621.60
2006	365	276.21	67.78	173.81	259.99	480.39
2007	365	250.52	26.36	196.70	246.84	317.02
2008	365	255.49	52.21	156.97	252.03	412.47
2009	365	216.10	31.46	169.89	206.72	330.58
2010	365	226.22	25.03	156.54	231.27	292.37
2011	365	243.57	20.14	193.75	245.99	285.22
2012	365	253.13	33.49	181.47	247.17	349.43
2013	365	189.74	32.30	139.57	182.08	303.46
2014	365	221.51	35.01	156.21	219.43	315.61
2015	365	249.39	28.98	194.92	251.31	319.26
2016	365	266.84	42.48	178.64	273.85	361.87
2017	365	220.26	17.69	178.45	222.40	256.07
2018	151	186.38	17.18	150.25	189.18	231.42