

Disentangling Demand and Supply of Media Bias: The Case of Newspaper Homepages

Tin Cheuk Leung, Koleman Strumpf

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Disentangling Demand and Supply of Media Bias: The Case of Newspaper Homepages

Abstract

In this study, we propose a novel approach to detect supply-side media bias, independent of external factors like ownership or editors' ideological leanings. Analyzing over 100,000 articles from The New York Times (NYT) and The Wall Street Journal (WSJ), complemented by data from 22 million tweets, we assess the factors influencing article duration on their digital homepages. By flexibly controlling for demand-side preferences, we attribute extended homepage presence of ideologically slanted articles to supply-side biases. Utilizing a machine learning model, we assign "pro-Democrat" scores to articles, revealing that both tweets count and ideological orientation significantly impact homepage longevity. Our findings show that liberal articles tend to remain longer on the NYT homepage, while conservative ones persist on the WSJ. Further analysis into articles' transition to print and podcasts suggests that increased competition may reduce media bias, indicating a potential direction for future theoretical exploration.

JEL-Codes: D220, D720, D830, L820.

Keywords: media bias, media economics, social media, machine learning.

Tin Cheuk Leung
Department of Economics
Wake Forest University
Winston-Salem / NC / USA
leungtc@wfu.edu

Koleman Strumpf
Department of Economics
Wake Forest University
Winston-Salem / NC / USA
strumpks@wfu.edu

January 2024

We would like to thank seminar participants from APIOS, CREED, CUHK, HKEA Biennial Conference, SEA, CESifo Conference on Economics of Digitization for their comments, Sean Anderson of the WFU HPC Facility for his help on coding, and Tom Cleveland on sharing his code for the NYT scraper. The usual disclaimer applies.

1 Introduction

Media bias remains a significant concern in the United States, as evidenced by extensive scholarly documentation and the consequent low levels of public trust in the media.¹ Adding to this discourse, Bennet (2023) provides a critical examination of bias within a leading publication like the New York Times (NYT). As a former NYT op-ed editor, Bennet (2023) delves into the shifting paradigms of bias at the NYT, suggesting two potential drivers: a demand-driven shift due to the newspaper’s evolving business model from advertising towards subscription revenue, and a supply-driven change reflecting a transformation in the reporters’ ideals, with a diminished focus on objectivity. This dual perspective adds depth to the exploration of media bias, underscoring the need to discern whether it originates from consumer demand or journalistic supply. Understanding the sources of this bias is vital, especially in light of the ambiguous effects competition may have on it. The theoretical literature, summarized by Gentzkow, Shapiro, and Stone (2015), posits a nuanced relationship: “When supply-side incentives lead firms to distort their reports, competition tends to reduce bias and enhance welfare. Conversely, when consumers demand bias, competition might intensify bias to cater to these preferences.”

This paper presents a new approach to analyzing media bias, focusing on supply-side factors without needing information about a newspaper’s ownership or editorial preferences.² Our method examines how long articles stay on a newspaper’s digital homepage, using this duration along with reader engagement data to assess bias. By controlling for demand-side influences on article duration, we can attribute any prolonged presence of articles with specific ideological leanings to supply-side preferences. This approach is adaptable to any newspaper, relying solely on data on homepage duration and reader engagement. To illuminate our analysis, let us explore two paradigmatic cases from NYT.

The first, “For Many Who Marched, Jan. 6 Was Only the Beginning” (Dias and Healy, 2022), published on January 23, 2022, profiles individuals who participated in the

¹Brenan (2022) finds that only 34% of Americans view the mass media as reporting news “fully, accurately, and fairly”, with a mere 7% expressing substantial trust and confidence, and 27% acknowledging a moderate amount. Additionally, Shearer (2020) highlights that two-thirds of U.S. adults perceive a bias in news reporting, often skewing towards a particular political stance.

²In this paper, we follow Gentzkow and Shapiro (2010) and Gentzkow, Shapiro, and Stone (2015) to define bias as “systematic differences in the mapping from facts to news reports— that is, differences which tend to sway naive readers to the right or left on political issues.”

Jan. 6 march. The article, marked by a pro-Democratic slant, describes the incident as a “dark day for the nation” and the “worst attack on American democracy”. Despite its early morning publication at 3am, the piece was featured on the NYT homepage by 6am, amassed approximately 200 tweet shares on the internet by noon on the same day, and sustained its position for over two days.

Contrastingly, “At Least 46 Migrants Found Dead in Tractor-Trailer in San Antonio” (Dobbins, Goodman, and Sandoval, 2022), published on June 27, 2022, endured a different trajectory. This article chronicles the harrowing death of 46 migrants in Texas, highlighting Gov. Greg Abbott’s statement—a Republican—who directly attributes the incident to President Biden’s policies: “These deaths are on Biden. They are a result of his deadly open border policies. They show the deadly consequences of his refusal to enforce the law.” This piece, published at 9pm, immediately featured on the homepage, and had more than 600 tweet shares within two hours, but was curiously removed within the next hour.

These two disparate editorial decisions provide an intriguing gateway into the exploration of media bias, inviting a comprehensive examination of both demand-side consumer preferences and supply-side editorial inclinations.

This study employs a novel dataset of articles appearing on the homepages of the NYT and Wall Street Journal (WSJ), collected from August 2021 to May 2023 for the former and from October 2022 to May 2023 for the latter. The dataset comprises over 80,000 NYT articles and more than 20,000 WSJ articles, supplemented by Twitter data gathered via the Twitter API for each article in the sample, serving as a proxy for article popularity (reader engagement or consumption, a connection we document in our analysis). The NYT sample boasts approximately 22 million tweets (averaging around 275 tweets per article), while the WSJ sample registers close to 2 million tweets (averaging about 100 tweets per article).

To calculate the sentiment and political leanings of the articles, we employ textual analysis to generate sentiment scores and pro-Democrat scores. The pro-Democrat scores, gauging an article’s political slant, were constructed via a machine-learning approach utilizing a training set of tweets from roughly 2000 American politicians, annotated with their political affiliations. An examination of the summary statistics reveals a discernible liberal bias in NYT articles (average pro-Democrat score of 0.6 on a scale where 1 signifies complete pro-Democrat alignment and 0 suggests absolute pro-Republican alignment)

whereas WSJ articles displayed more neutrality (average score of 0.53).

We first investigate the relationship between an article’s presence on the homepage of major newspapers and its ensuing online visibility, as measured through tweet counts. Using the exogenous updating of homepage as an instrument for homepage appearance, we find a marked enhancement in tweet counts for articles featured on the homepage. Specifically, for the NYT, homepage presence leads to a 35% rise in tweets, with the effect varying across types: op-ed articles witness a 76% surge, while feature articles see a more modest 26% increase. The effect is more pronounced for the WSJ, where articles benefit from a 162% uptick on average: news articles gain a 150% boost and feature articles experience a robust over 200% uplift.

Our core analysis adopts a parametric survival model to investigate the length of time an article remains on the homepage of the respective newspaper websites. The findings underscore the significant role of demand-side factors, particularly the impact of hourly tweets on article hazard rate, indicating that articles accruing more tweets in an hour are associated with a lower hazard rate.

Supply-side factors are also influential, especially in the case of NYT where more liberally skewed articles tend to have a longer homepage presence, even after controlling for demand-side factors. Conversely, for WSJ, supply-side factors generally display statistical insignificance. However, when focusing on specific sections (e.g., US) or articles with certain politically charged keywords (such as “Politics”), more conservative articles have longer homepage survival times, even after adjusting for demand-side factors.

As noted before, if media bias is driven mainly by supply such as the ideological leaning of journalists and editors, then theoretically a more competitive media market is likely to mitigate the degree of media bias. Central to this argument is the principle that increased competition exposes audiences to diverse perspectives, subsequently necessitating accuracy and reducing scope for one-sided narratives. Building on this theoretical foundation, an emergent hypothesis is the interplay between the elasticity of demand and media bias. Specifically, when demand is more elastic, the propensity for media bias might be diminished, given that media outlets would be vying for a broader, more heterogeneous audience base, thereby minimizing any inherent bias.

Our empirical endeavors aimed to test this hypothesis, contrasting two distinct platforms: the print edition of NYT and both NYT and WSJ’s flagship podcasts. The print medium, traditionally understood to cater to a stable and consistent readership, can be characterized

by a more inelastic demand. Conversely, podcasts, with their rapidly expanding and varied listenership, can be perceived as operating in an environment with more elastic demand. Our findings from the print edition analysis for the NYT sample revealed stronger biases, particularly within specific sections, suggesting that when demand is relatively inelastic, supply-side factors exert more influence on editorial decisions.

In stark contrast, our analysis of podcast editorial decisions for both NYT and WSJ unveiled a different landscape. Demand-side factors, such as the popularity of an article measured by tweet counts, emerged as influential determinants. Supply-side factors, notably the pro-Democrat scores, manifested no substantial correlation with podcast feature decisions. This underscores the literature’s assertion that in contexts of heightened competition and more elastic demand, media bias is potentially attenuated.

In summation, our analyses of both the print and podcast platforms resonate with the theoretical insights of the literature.

The remainder of the paper is as follows. Section 2 reviews the related literature, and Section 3 discusses the importance of homepage to news media websites. We describe the data in Section 4. We then discuss our main analysis and present our empirical results in Sections 5 and 6. Section 7 presents our results on the print edition and podcast analysis. The last section concludes.

2 Literature

This paper contributes to the study of media bias. Groseclose and Milyo (2005) are among the first studies to systematically document media bias in a large scale. They find that all but two (*Fox News’ Special Report* and the *Washington Times*) news outlets received scores to the left of the average member of Congress, while NYT received scores far to the left of center. Many studies have since analyzed several drivers of media bias,³ including demand for media bias (Huang, Meng, and Weng, 2022; Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017, Puglisi and Snyder, 2011), supply of media bias (Ansolabehere, Lessem, and Snyder, 2006; Baron, 2006; Cagé et al., 2022; Demsetz and Lehn, 1985; Martin and McCrain, 2019; Strömberg, 2004), professional norms (Baum and Groeling, 2008; Shapiro, 2016), journalists’ ideology (Boxell and Conway, 2022), and

³Gentzkow, Shapiro, and Stone (2015) and Puglisi and Snyder (2015) review the theoretical and empirical literature on media bias, respectively.

the level of market competition (Chan and Suen, 2008; Mullainathan and Shleifer, 2005; Gentzkow and Shapiro, 2008; Qin, Strömberg, and Wu, 2018). Several papers also study the historical evolution of partisan news sources (Gentzkow, Glaeser, and Goldin, 2006; Hirano and Snyder, 2020) and the importance of fake news and partisan news in a typical American’s news diet (Allen et al., 2020; Muise et al., 2022). While many of the empirical papers above analyzes local newspapers, our paper is one of the first to study the supply factors of media bias in major national newspapers.

Our paper contributes to the growing literature of the relationship between social media and news. Several papers study the competition for attention between newspapers and social media (de Cornière and Sarvary, 2022; Jeon and Nasr, 2016), the impact of social media on fake news (Allcott and Gentzkow, 2017), and the impact of social media on news production by mainstream media (Cagé, Hervé, and Mazoyer, 2022; Hatte, Madinier, and Zhuravskaya, 2022). Our paper contributes to the literature by using tweets as a proxy of an article’s popularity and therefore showing that demand side factors play a role in newspapers’ homepage decision.

This paper also contributes to our understanding of the impact of technology in general and in the media industry. Athey and Luca (2019) documents the use of various statistical tools made possible by digital technology, including A/B testing, in Tech companies. Previous studies have examined the impact of technology/statistical tools in the media industry such as how A/B testing algorithms are changing workflow and headline writing practices (Hagar and Diakopoulos, 2019) and also articles’ performances (Leung and Strumpf, 2023), the returns to data and informational externalities associated with algorithmic recommendations relative to human curation in online news (Claussen, Peukert, and Sen, 2023), the impact of news aggregator on media companies’ revenue and concentration (Calzada and Gil, 2020; Calzada, Duch-Brown, and Gil, 2022), how ad-avoidance technologies might increase advertising clutter and reduce content quality (Anderson and Gans, 2011), and how NYT expansion of its home delivery affects local newspaper circulation (George and Waldfogel, 2006). Our results shed light on the use and the impact of digital technology (digital homepage) on media companies decision to promote articles and to influence media bias.

Lastly, this paper is related to the literature on digital platform bias. Several theoretical papers explore the reasons behind why platforms might bias the rankings of products (Armstrong and Zhou, 2011; Hagiu and Jullien, 2014; Parker, Petropoulos, and Van Alstyne,

2020; de Cornière and Taylor, 2019; Bourreau and Gaudin, 2022). Some empirical papers attempt to test for platform bias, including demonstrating bias in Amazon’s “frequently bought together” recommendations or Amazon’s buy box (Chen and Tsai, 2023; Raval, 2023) or measuring bias in search ranking directly (Jürgensmeier and Skiera, 2023; Farronato, Fradkin, and MacKay, 2023; Aguiar, Waldfogel, and Waldfogel, 2021; Reimers and Waldfogel, 2023).

3 The Importance of Newspaper Homepage

The homepage of a newspaper website stands as its most significant digital real estate, serving as the primary gateway through which visitors access the plethora of articles and multimedia content. A Pew Research Center study in 2011 underscores this centrality, highlighting that for a majority (21 out of 25) of the sites they examined, including NYT and WSJ, the homepage emerged as the most frequented section (Olmstead, Mitchell, and Rosenstiel, 2011).

Fast-forwarding to more contemporary data, the pivotal role of the homepage in drawing readers becomes even more palpable. The 2022 annual report from the New York Times Company reveals staggering figures regarding their digital reach: approximately 99 million unique visitors from the United States alone accessed NYTimes.com monthly through desktop or mobile devices. This number swells to around 145 million unique visitors when considered on a global scale, encompassing readership from every corner of the globe (New York Times, 2022).

A newspaper homepage is typically the first place a reader views, and so shapes which articles he eventually reads. Table 1, using data from SimilarWeb Pro between May and June 2023, shows that over 50% of total traffic (desktop and mobile) for each of the NYT and WSJ is direct traffic. Direct traffic is when the user starts reading by opening the newspaper app or enters a url or bookmark into their browser, and in almost all these cases means the user is starting on the paper’s homepage. This is a conservative measure of the homepage’s importance since some of the other traffic may also start on the homepage (this might include referrals, paid or organic search, or social media). Such statistics indicate that a significant majority of visitors are likely accessing the site directly, presumably via the homepage, rather than through secondary channels or external referrals. This understanding not only reaffirms the homepage’s unmatched value in the digital realm

of news but also emphasizes the profound implications of editorial decisions about what content to spotlight there.

Table 1: % of Direct Traffic to NYT and WSJ Websites

	NYT		WSJ	
	US	World	US	World
Mobile	48.4%	46.7%	51.5%	49.9%
Desktop	59.8%	57.9%	63.4%	61.3%
Total	52.3%	50.4%	56.3%	54.6%

Data: SimilarWeb Pro (May-June 2023)

4 Data

Most of our data come from Leung and Strumpf (2023). In particular, we use a scraper to collect our main news articles data from the NYT and WSJ, and complement the data with tweets data from the Twitter API.

4.1 NYT and WSJ Data

We have a scraper which collects data from the NYT homepage, <https://www.nytimes.com>, and ran every minue from August 2021 through June 2023. On each cycle it collects all articles on the NYT homepage, including the headline, url, and article id. We also use the official NYT API, <https://developer.nytimes.com/apis>, to collect information on every NYT article (including those not on the homepage). The API returns each article’s metadata, including article id, print headline (if any), when the article was first posted online, news tone,⁴ section, and a subset of the article text (abstract and first paragraph) for all articles published in the sample period (whether they were ever on the homepage or not). We also obtained the top 20 articles in terms of email, share, and views for each hour in the sample period. Overall, there were 81,692 articles published by the NYT in the sample period, 48,501 (59%) of these were put on the homepage at some point.

⁴The three news tones are “News” (which is mainly hard news), “Opinion” (which is mainly op-ed), and “Feature” (which is mainly soft news).

For the WSJ, we started to scrape the WSJ homepage, <https://www.wsj.com/>, on Oct 8, 2022, in a similar manner as with the NYT. In particular the Python scraper uses a randomly generated user-agent and does not accept cookies. Similar to the NYT scraper, the WSJ scraper collects all articles on the WSJ homepage, including the headline, url, and article id. Because the WSJ lacks an API, we scraped the WSJ archive, <https://www.wsj.com/news/archive/years> to obtain the urls of all WSJ articles in the sample period and then used a Python package, *newspaper3k*, to extract articles’ metadata, including article id, when the article was first posted online, section, and a subset of the article text (abstract and first paragraph) for these articles. As with the NYT API, the WSJ archive includes articles never reaching the homepage.

There are several things worth noting. First, our WSJ homepage scraper was not able to scrape the op-ed articles. Second, we do not have information on print headline and news tone as in the NYT data. Instead, we categorize an article’s news tone by its section which aligns with the NYT sorting.⁵

Overall, there were 20,607 articles published by WSJ in the sample period, 69% (14,250) of these were put on the homepage at some point.

4.1.1 Textual Analysis

We follow Leung and Strumpf (2023) to use textual analysis to construct two measures for the abstracts (and tweets in later sections) of an article.⁶ The first measure is a sentiment score of the text. In particular, we use the VADER sentiment analysis tool (Hutto and Gilbert, 2014) of the Natural Language Toolkit (NLTK) in Python (Bird, Klein, and Loper, 2009) to implement sentiment analysis for the abstracts of each article in the sample. As Hutto and Gilbert (2014) shows, VADER’s model incorporates syntax and punctuation rules, and is validated with human coding, making its sentence prediction 55–96% accurate, which is on par with Stanford Sentiment Treebank, a method that incorporates a more complex computational algorithm. Also, VADER includes the word banks of established tools like LIWC, ANEW, and GI, as well as special characters such as emoticons and cultural acronyms (e.g., LOL), which makes it particularly suited for social

⁵The “News” tone includes sections including “Business”, “Economy”, “Markets”, “News”, “Politics”, “Science”, “US”, and “World”. The “Feature” tone includes sections including “Arts”, “Life”, “Style News”, “Style and Substance”, “WSJ Puzzles”, “Tech”, “Real Estate”, and “Magazine”.

⁶Interested readers are referred to the paper for more details on the construction of these measures.

media jargon.⁷ In addition to social media text, Hutto and Gilbert (2014) also shows that VADER performs well across several domain contexts including NYT editorials, movie reviews, and product reviews.

The second measure is a measure of political slant. We follow Boxell and Conway (2022) and Leung and Strumpf (2023) to adopt a machine learning approach to construct our measure of political slant. To construct our training and testing data, we first obtain the Twitter accounts’ information of approximately 2,100 politicians (with known affiliations with either the Democratic Party or the Republican Party) from ProPublica, an investigative journalism website.⁸ We then use the Twitter API to obtain all their tweets (approximately 2 million tweets in total).

After applying the standard pre-processing steps, we then use the Linear Support Vector Classification (Linear SVC) model in the scikit-learn package, a Python module, as our machine learning model. We test the model using the training dataset (approximately 1.4 million tweets). Overall, the model achieves 82% accuracy on the leave-out validation dataset. In addition to this, we also construct the confusion matrix, which shows the the percentage of correct and incorrect classifications based on each class (Democrat or Republican), to evaluate the goodness of fit of the model. Figure 1 shows that our model can predict a Democrat’s tweet correctly 84% of the time, with only 16% chance being wrongly labeled as a Republican’s tweet. It can also predict a Republican’s tweet correctly 80% of the time, with 20% chance being wrongly labeled as a Democrat’s tweet.

We apply the model to “predict” whether the text in our main dataset is written by a Democrat politician. The confidence of the prediction (in terms of probability of it being written by a Democrat) is our measure of slant.

Figure 2 presents the lexical trends within U.S. news articles from the NYT and WSJ, several distinct patterns emerge.⁹ The wordclouds are bifurcated into two main panels: the upper panel showcases the linguistic trends in NYT articles, with the left side representing pro-Republican articles (articles with pro-Democrat scores below 0.3) and the right side indicating pro-Democrat ones (articles with pro-Democrat scores above 0.7). Conversely, the lower panel elucidates the most frequent terms within WSJ articles, adhering to the same political score distinctions.

⁷We also use VADER to obtain sentiment scores for tweets which will be discussed in Section 4.2.

⁸We dropped the independent politicians in our sample.

⁹We present more wordclouds in other news sections in Appendix A.

Figure 1: Confusion Matrix of Machine Learning Prediction

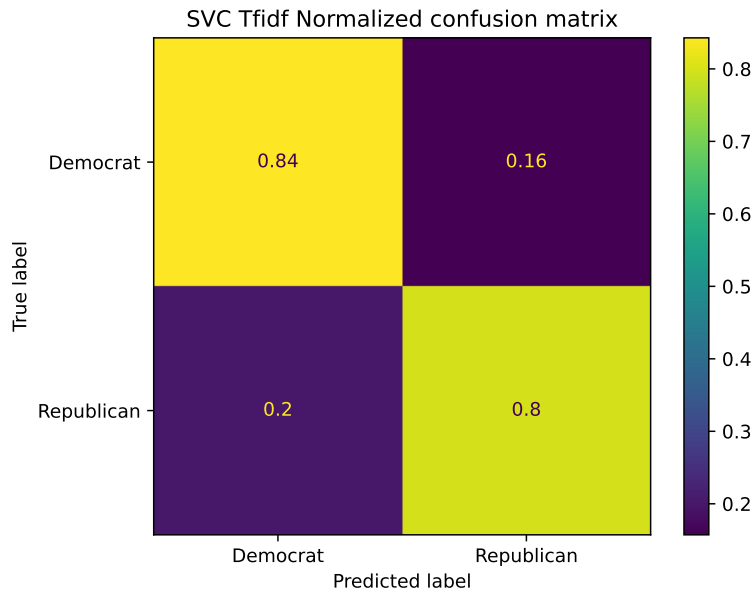
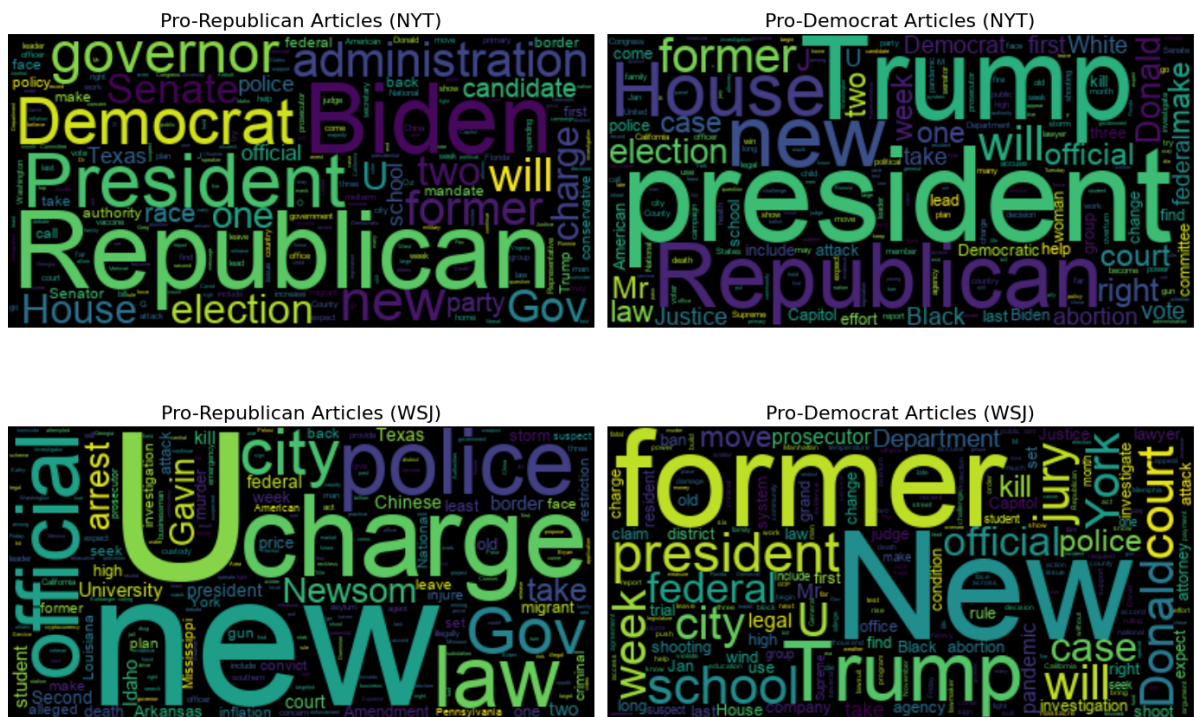


Figure 2: Word Clouds of US News Articles' Abstracts



Both the NYT and WSJ’s pro-Democrat articles manifest a notable overlap in frequently used terms. Common lexemes such as “former”, “President”, “Trump”, “abortion”, “Black”, and “shooting” underscore this shared thematic emphasis.

However, when comparing the vocabulary employed in pro-Republican articles between the two publications, discrepancies become apparent. In the NYT, pro-Republican articles predominantly feature terms such as “Republican”, “President”, “Biden”, and “Democrat”. Contrastingly, WSJ’s pro-Republican discourse gravitates towards different terms, notably including “police”, “Gavin”, “Newsom”, and “border”.¹⁰ These variations hint at differing editorial foci and narrative frameworks adopted by the two newspapers.

4.1.2 Summary Statistics

An integral component of understanding media bias involves analyzing the duration for which different types of articles remain on a media outlet’s homepage. Figure 3 provides a comprehensive overview of this aspect, presenting the distribution of homepage duration for articles from NYT and WSJ.

In the left panel of Figure 3, we observe the distribution for NYT articles. In comparison, the right panel shows the same for WSJ. A striking observation from this analysis is that a significant number of articles (5% for NYT and 8% for WSJ) were taken off the homepage within the initial 2 hours of their placement.

Delving further into the types of content, it is noteworthy that op-ed articles in NYT tend to have a longer presence on the homepage compared to other categories of articles. The bottom right panel of the NYT subgraph shows that more than 15% of op-ed articles stay on for more than 48 hours, whereas less than 3% of articles of other news tones (for both NYT and WSJ) stay on the homepage for more than 2 days. This implies that editorial content could potentially exert a more substantial influence on the readers due to its longer visibility duration.

Large spikes are also noticeable when we examine the timeline of content removal from the homepage. Editors in both NYT and WSJ appear to withdraw more articles after specific time intervals post-placement, such as 6 or 12 hours, which is especially true for Feature articles in NYT. This finding suggests that there could be editorial strategies or guidelines in place to refresh the homepage content at regular intervals, affecting the

¹⁰There are a few terms such as “Texas” which feature on both NYT and WSJ’s pro-Republican articles.

overall exposure of different news stories.

Figure 3: Histogram of Homepage Duration by News Tone



Table 2 summarizes the articles' statistics for the two newspapers, including the sentiments conveyed in their headlines and abstracts, and the political slant of their content.

Table 2: Summary of News Articles

	NYT		WSJ	
	Homepage (No)	Homepage (Yes)	Homepage (No)	Homepage (Yes)
Sentiment Scores (Abstract)	0.0690 (0.392)	-0.0140 (0.454)	0.0348 (0.387)	0.0161 (0.423)
Pro-Dem. Scores (Abstract)	0.603 (0.205)	0.598 (0.215)	0.519 (0.213)	0.539 (0.209)
tone (News)	0.517 (0.500)	0.614 (0.487)	0.193 (0.395)	0.643 (0.479)
tone (Opinion)	0.00925 (0.0957)	0.133 (0.340)		
tone (Feature)	0.471 (0.499)	0.253 (0.435)	0.250 (0.433)	0.231 (0.422)
N	33191	48501	6357	14250

Mean coefficients; SD in parentheses

Firstly, it is noteworthy that between 60% and 70% of all articles from both the NYT and WSJ have appeared on their respective homepages at some point. This reveals a substantial proportion of articles deemed significant enough by the editors to warrant homepage exposure.

Secondly, the sentiment analysis indicates a mild preference for slightly more negative tones in the abstracts of articles that have graced the homepage, as opposed to those that have not. This observation holds for both the NYT and WSJ, suggesting a potential correlation between negative sentiment and the perceived newsworthiness of an article.

Thirdly, the political slant of the articles offers fascinating insights. The NYT articles' political slant scores, at approximately 0.6, are more pro-Democrat or liberal compared to the WSJ, which exhibits scores between 0.52 and 0.54. These findings corroborate the commonly perceived political inclinations of these two major publications.

Lastly, when comparing articles that have appeared on the homepage to those that have not, the difference in political slant within the same publication is slight for the NYT, but more evident for the WSJ. For WSJ, articles that have appeared on the homepage tend to bear a slightly more liberal slant than those that have not.

4.2 Twitter API

Ideally, we would measure the popularity of articles with the actual views and share data from NYT. Unfortunately, we do not have access to it. Instead, we use the number of tweets that link an article's URL as our measure of its popularity (reader demand).

News plays a prominent role on Twitter. In a Pew Research Center study that surveyed 2,548 Twitter users in 2021, Anderson, Rainie, and Nolan (2021) reveal that 23% of Americans use Twitter, and roughly seven-in-ten US Twitter users (69%) say they get news on the site. While only 27% of Americans report to have at least some trust in the information they find on social media in general, two-thirds of Twitter news consumers say they have at least some trust in it. Twitter users' reliance on Twitter as a news source has also grown over time. In the same study, the authors report that "fully 70% of Twitter news consumers say they have used Twitter to follow live news events, up from 59% who said this in 2015."

In January 2021, Twitter opened its full tweet archive to academic researchers. We use the Twitter API to extract all tweets containing the URL links of all the NYT and WSJ articles in our sample. For each tweet, we extract the tweet content, the date and time of the tweet, the number of likes, the number of retweets, and other information of the authors such as the name, the Twitter id, and their self-reported locations. We also construct a sentiment scores for each tweet in the same way as the abstracts. We end

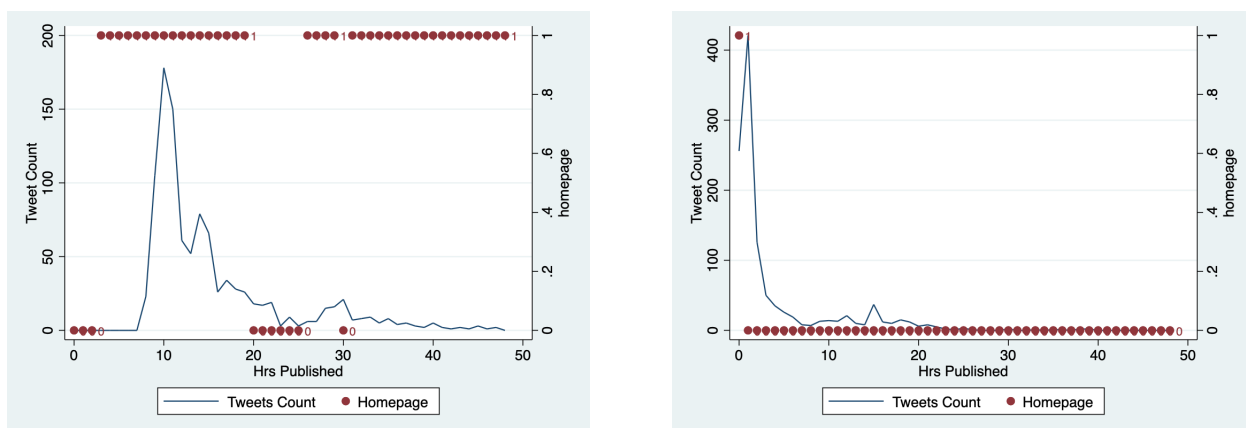
up with a database of over 24 million tweets (including retweets) with 22 million tweets for the NYT sample (approximately 275 tweets per article), and 2 million tweets for the WSJ sample (approximately 100 tweets per article).

Figure 4 provides a compelling illustration of how the volume of social media activity (measured in terms of tweet counts) interacts with an article’s presence on a newspaper’s homepage. The two case studies analyzed in this figure are the articles discussed in the introduction - ”For Many Who Marched, Jan 6 Was Only the Beginning” and ”At Least 46 Migrants Found Dead in Tractor-Trailer in San Antonio”.

Figure 4: A Tale of Two NYT Articles

”For Many Who Marched, Jan 6 Was Only the Beginning”

”At Least 46 Migrants Found Dead in Tractor-Trailer in San Antonio”



The left panel of the figure corresponds to the first article, which adopts a more liberal tone (the pro-Democrat score is 0.93). The article was initially published at 3am and placed on the NYT homepage at 6am. Interestingly, tweet counts did not show any significant increase until several hours after the article was featured on the homepage. The peak hourly tweet count reached was 178. Except for a brief period in the following night, the article was kept on the homepage for more than two days. By the end of this 48-hour window, the tweet count had decreased to zero, indicating that the public interest, as gauged by Twitter activity, waned over time.

On the other hand, the right panel of the figure depicts the fate of the second article, which is less liberal in its tone (the pro-Democrat score is 0.22). The article was published at 9pm and was immediately placed on the homepage. The tweets count in the first hour following publication already reached 252, surpassing the peak tweets count of the first

article. Despite this initial surge in social media activity, the article was removed from the homepage within the next hour, even as the tweet count increased further to more than 400 in the subsequent hour.

These contrasting trajectories of the two articles hint at supply-side factors that influence media bias, specifically in terms of when a newspaper decides to remove an article from its homepage. While demand-side factors, like the level of social media activity, play a role, they do not fully account for the observed patterns. Therefore, it is crucial to further investigate the potential impact of the media outlet’s inherent political inclinations on such editorial decisions.

4.2.1 Summary Statistics

The data compiled in Table 3 presents a compelling case for using the number of tweets as an effective proxy to gauge the popularity of articles. The table highlights the stark difference in hourly tweet count between articles that secured a position in the top 20 as compared to those that did not across all rankings (views, shares, and emails).

Table 3: NYT Articles’ Rank and Hourly Tweets Count

	Views	Share	Email
Outside Top 20	4.894 (46.66)	4.538 (40.90)	5.007 (45.53)
Rank 11-20	13.20 (71.65)	16.03 (114.8)	12.33 (92.24)
Rank Top 10	14.96 (66.20)	24.32 (48.25)	17.01 (87.11)

Mean coefficients; SD in parentheses

For articles that did not achieve top 20 status, the hourly tweets count ranges between 4.5 to 5, a figure considerably lower than their top 20 counterparts. Articles making it to the top 20 exhibit a much higher hourly tweets count, ranging between 12 to 24 on average.

A deeper dive into the data reveals a further discrepancy within the top 20 group: articles that secured a position within the top 10 enjoy a higher tweet count on average than articles ranked between 11 and 20. The disparity is most pronounced when examining share rankings, where the top 10 articles accrue an average of 24 tweets per hour, compared

to 16 tweets for those in the 11-20 band. A similar, albeit less marked, trend is observable in the email rankings, with the top 10 averaging 17 tweets per hour and positions 11-20 garnering 12 tweets on average.

In summary, the findings substantiate the reliability of using tweet counts as an effective proxy to estimate an article’s popularity, thus lending credibility to its application in our analysis.

Table 4 offers a quantitative look into the Twitter engagement for articles from the two newspapers. Several key points stand out from this analysis.

Table 4: Summary of News Articles’ Tweets Count

	NYT		WSJ	
	Homepage (No)	Homepage (Yes)	Homepage (No)	Homepage (Yes)
Tweets Count	63.46 (537.2)	404.3 (1501.7)	46.06 (225.8)	138.8 (1032.7)
Tweets Count (Neg)	5.339 (36.88)	35.72 (160.5)	4.175 (21.16)	12.22 (88.53)
Tweets Count (Pos)	5.313 (74.36)	40.42 (207.6)	4.694 (30.69)	13.01 (117.8)
Tweets Count (Pro-Dem.)	39.21 (371.1)	244.0 (922.4)	19.57 (80.45)	64.60 (355.5)
N	33191	48501	6357	14250

Mean coefficients; SD in parentheses

First, the data reveals that articles from the NYT are more than twice as popular as those from the WSJ in terms of Twitter engagement, with average tweet counts standing at 266 and 110, respectively. When narrowing our focus to articles that were featured on their respective newspaper’s homepage, the popularity gap widens further. In this context, NYT articles receive roughly three times as many tweets (404 on average) as their WSJ counterparts (139 on average).

Second, articles that have appeared on the homepage garner significantly more Twitter engagement for both publications. NYT articles featured on the homepage attract about six times as many tweets, while WSJ articles enjoy a threefold increase in tweets. This might suggest that articles which secure a homepage position are inherently more newsworthy or interesting, triggering greater engagement. Alternatively, it could imply that the homepage placement itself amplifies an article’s visibility and hence its popularity—or both factors may be at play.

Third, approximately 20% of tweets relating to articles from both newspapers express a

discernible sentiment—either positive or negative—with a slight tilt towards the positive. This fact points to the relatively balanced sentiment distribution in readers’ responses to these articles.

Finally, the political leanings of tweets connected to NYT and WSJ articles also differ. Approximately 60% of tweets related to NYT articles lean liberal, compared to 47% for WSJ articles. This finding aligns with the previously observed political slants of these two publications.

5 Impact of Homepage Presence on Tweets Count

Before delving into the intricate interplay between supply and demand forces in shaping media bias, we need to ask the question: Why should we concern ourselves with the editorial decisions dictating when articles are removed from the homepage? The homepage, often seen as the digital embodiment of a newspaper’s front page, holds paramount importance in shaping readers’ perceptions and directing their attention. Editorial decisions about which stories linger and which are withdrawn not only reflect inherent biases but also significantly influence the dissemination and reception of news.

In the quest to ascertain the influence of homepage presence on the popularity of articles, as gauged by tweets count, we confront several methodological challenges. Intuitively, one might consider running a straightforward regression with tweets count as the dependent variable and homepage presence as an independent variable. However, as elucidated in our preceding analyses, such an approach is plagued by the endogeneity of homepage presence. This is because articles do not find their way to the homepage randomly; they are strategically selected based on various unobserved factors, potentially confounding the relationship between homepage presence and article popularity.

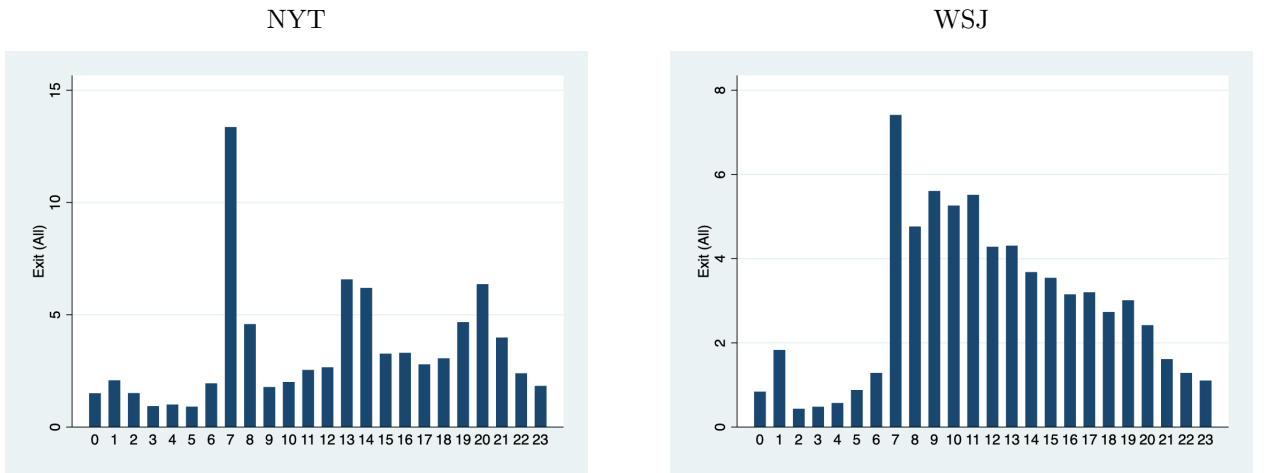
We address this issue of endogeneity with an instrumental variable (IV) approach. Our chosen IV emerges from a pattern discerned in the temporal dynamics of article withdrawal from the homepage.

As depicted in Figure 5, the average number of articles removed from the homepage varies distinctly by the hour. This bifurcation, illustrated separately for the NYT on the left panel and the WSJ on the right, reveals certain hours of the day where newspapers seem to undergo more pronounced homepage updates. These updates, importantly, appear to be driven by routine editorial practices rather than the inherent demand for

specific articles.

For the NYT, salient spikes in article withdrawal are observed at 7am, 8am, 1pm, 2pm, 7pm, and 8pm. For the WSJ, the spikes are more concentrated, primarily at 7am and 1am. Leveraging these temporal patterns, we deploy these specific hours as instrumental variables in our analysis. By doing so, we aim to circumvent the endogeneity issue and uncover a more accurate estimation of the impact of homepage presence on article popularity, as proxied by tweets count.

Figure 5: Articles Withdrawal by Hours of Day



In order to examine the impact of homepage presence on the popularity of articles as measured by tweet counts, we employ the following regression specification:

$$\ln Y_{it} = \beta_0^H \text{Homepage}_{it} + \beta_1^H X_{it} + \phi_i + \varepsilon_{it}, \quad (1)$$

where $\ln Y_{it}$ represents the natural logarithm of the zero-inflated tweet counts pertaining to article i at time t .

Our primary independent variable of interest, Homepage_{it} , is a binary indicator capturing whether article i is prominently displayed on the homepage at time t . This serves as our key measure to understand the immediate influence of homepage placement on the virality of articles as gauged by tweet counts. Recognizing the potential endogeneity of Homepage_{it} , which could confound our estimates, we employ an instrumental variable approach. Specifically, we use $\text{WithdrawalHours}_{it}$ as an instrument, which is a binary indicator taking the value of one if time t coincides with the specific hours during which

the newspaper partakes in its routine editorial practice of refreshing its homepage. This instrumental strategy allows us to exploit variations induced by these editorial practices to tease out the causal effect of homepage presence on tweet counts.

The vector X_{it} encompasses a set of controls that account for various determinants of article popularity. This includes the lagged logarithm of the zero-inflated tweet counts, denoted as $\ln Y_{it-1}$, to control for the article’s inherent virality in the preceding time period. Moreover, we adjust for the number of hours since the article’s publication to account for potential decay effects in attention and visibility over time. To further refine our estimations, we introduce a series of time fixed effects. These entail controls for specific times of the day, days of the week, and months, capturing cyclical patterns in readership and online activity.

Lastly, to comprehensively account for any persistent differences across articles that might be influencing their tweet counts, we incorporate article-specific fixed effects, ϕ_i . These effects mitigate biases arising from unobserved heterogeneity across articles that remain constant over time.

Our findings, illustrated in Table 5, present a comprehensive analysis of the repercussions of homepage placement on tweet counts for articles from both newspapers. For uniformity and consistency in our analysis in Section 6, our sample solely includes articles which were on the homepage in the preceding hour, $t - 1$. This criterion aligns our sample with the forthcoming investigation on article longevity on the homepage.

Table 5: Ln Tweets Regression with Article F.E.

	NYT				WSJ		
	News	Opinion	Feature	All	News	Feature	All
	Second Stage. Dependent variable is Ln Tweets						
Homepage	0.320*** (0.0386)	0.764*** (0.171)	0.263*** (0.0215)	0.349*** (0.0241)	1.496*** (0.144)	2.070*** (0.319)	1.622*** (0.131)
Ln Tweets (Lag)	0.407*** (0.00172)	0.413*** (0.00230)	0.270*** (0.00265)	0.400*** (0.00119)	0.292*** (0.00403)	0.213*** (0.00928)	0.278*** (0.00374)
Hours Published	-0.0268*** (0.000487)	-0.0178*** (0.000497)	-0.0162*** (0.000332)	-0.0207*** (0.000181)	-0.0112*** (0.00118)	-0.00392* (0.00203)	-0.00916*** (0.000996)
Observations	384512	223509	137959	745980	141749	46921	188670
Time-of-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	First Stage. Dependent variable is Homepage						
Withdrawal Hours	-0.0716*** (0.00109)	-0.0190*** (0.000807)	-0.170*** (0.00179)	-0.0770*** (0.000706)	-0.0497*** (0.00251)	-0.0432*** (0.00468)	-0.0482*** (0.00221)
Ln Tweets (Lag)	0.0235*** (0.000563)	0.00680*** (0.000474)	0.0230*** (0.00127)	0.0222*** (0.000394)	0.0189*** (0.000987)	0.0212*** (0.00204)	0.0199*** (0.000890)
Hours Published	-0.0116*** (0.0000743)	-0.00283*** (0.0000331)	-0.00983*** (0.000123)	-0.00629*** (0.0000382)	-0.00787*** (0.0000994)	-0.00616*** (0.000163)	-0.00734*** (0.0000846)
Time-of-Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
 * $p < .1$, ** $p < .05$, *** $p < .01$

Delineating the table, the left panel displays results corresponding to the NYT, while the right panel is devoted to the WSJ. Each subtable is bifurcated into two parts: the upper half portrays the second-stage results, whereas the first-stage findings are outlined in the lower segment.

Commencing with the first-stage results, both for NYT and WSJ, there's an observed higher propensity for article withdrawal during those specific withdrawal hours. This observation is congruent with our prior expectations, which were predicated upon the routine editorial practices of these newspapers.

The crux of our analysis, the second-stage results, underscore the salience of homepage presence in magnifying an article's reach, gauged via tweet counts. However, the magnitude of this impact exhibits stark differences between the two newspapers. For the NYT, an article's presence on the homepage bolsters its tweet counts by an average of 35%. This augmentation is most pronounced for op-ed articles, witnessing a surge of 76%, while feature articles experience a relatively modest increment of 26%. Contrastingly, the WSJ articles witness a more dramatic impetus from homepage placement. On average, being

on the homepage amplifies an article’s tweet counts by a staggering 162%. Disaggregating by article type, news articles see a 150% surge in tweets when featured on the homepage, while feature articles experience an even more pronounced spike, exceeding 200%.

One plausible rationale behind the relatively attenuated impact of homepage presence for NYT articles could be the newspaper’s efficacy in amplifying its articles through alternative promotional channels, notably platforms like Facebook and Twitter. This differential effectiveness in external promotion strategies could potentially account for the observed discrepancies in the relative benefits of homepage placement between the two newspapers.

6 Homepage Survival Analysis: Model and Results

Our main analysis leverages a survival model to understand the distribution of a particular “time to event”, which in our case, is the duration an article stays on the homepage of the newspaper’s website. In the analysis, we include both demand and supply side factors, represented by variables such as sentiment, political slant, tweet counts, and timing of other new articles, influence the duration an article stays on the homepage.

The model employed is a parametric survival model with a Weibull distribution. The Weibull model is chosen for its flexibility; it can take on a variety of shapes and thus accommodate a variety of underlying hazard rate patterns. Its hazard function, defined as the risk of removal at any given time, is expressed as:

$$h(t; X_{it}, \beta, \lambda, \alpha) = \alpha\lambda(t\lambda)^{\alpha-1} \exp(\beta X_{it}), \quad (2)$$

where t is the number of hours since publication, λ is a scale parameter, and $\alpha > 0$ is the shape parameter. For $\alpha > 1$, the hazard function is monotonically increasing, meaning the risk of removal grows over time. The variable X_{it} is a set of (potentially) time-varying controls for article i , including the natural log of zero-inflated tweets, sentiment and pro-Democrat scores of the article, and number of new articles published by the newspaper at time t .¹¹ Additionally, the model incorporates several fixed effects dummies for the time of the article’s publication (hour of the day, day of the week, and year-month).

¹¹We use the number of new articles of the same news tone when we restrict the sample to a certain news tone.

In this study, we are particularly interested in distinguishing the demand and supply side factors contributing to media bias, a delineation facilitated by our model’s setup.

Firstly, on the *demand side*, we capture the article’s popularity through the inclusion of tweet counts as a control. We posit that if demand-side factors, such as reader interest, are significant drivers of the duration an article remains on a homepage, articles that generate higher tweet counts, signaling greater reader engagement and interest, would have an extended presence on the homepage.

Secondly, on the *supply side*, we introduce measures of an article’s political slant, specifically pro-Democrat scores. Both NYT and WSJ have different levels of baseline political slant, with NYT being more liberal and WSJ being more conservative. However, within each newspaper, the relative political slant of individual articles can vary widely. By incorporating the pro-Democrat scores of articles as a control, we can examine whether articles with a particular political leaning are more likely to remain on the homepage for extended periods. If supply-side factors, like the political bias of the newspaper’s editorial board, play a significant role, we might expect articles aligning more closely with the newspaper’s overall political slant to have prolonged homepage durations.

Therefore, the inclusion of these controls enables us to disentangle the effects of demand and supply side factors on media bias within the digital news landscape.

6.1 Main Results

In presenting our survival analysis results as displayed in Table 6, we delineate our findings separately for the NYT and WSJ, as featured in the left and right panels respectively. The results are presented in terms of hazard ratios, which provides a measure of the effect of each covariate on the probability of an event occurring, in this case, the removal of an article from the newspaper homepage. A hazard ratio less than one indicates that as the covariate increases, the event becomes less likely, i.e., the article stays on the homepage longer.

Table 6: Survival Analysis of Homepage (Hazard Ratios)

	NYT				WSJ		
	News	Opinion	Feature	All	News	Feature	All
Ln Tweets	0.970*** (0.00411)	0.734*** (0.0114)	1.121*** (0.0106)	0.982*** (0.00367)	0.929*** (0.00963)	0.845*** (0.0175)	0.908*** (0.00823)
Sentiment Scores (Abstract)	0.950*** (0.0111)	0.935** (0.0310)	0.997 (0.0205)	0.956*** (0.00928)	0.872*** (0.0201)	0.974 (0.0385)	0.904*** (0.0175)
Pro-Dem. Scores (Abstract)	0.911*** (0.0225)	0.888* (0.0538)	0.874*** (0.0388)	0.897*** (0.0182)	0.986 (0.0472)	0.922 (0.0779)	0.948 (0.0386)
# of New Articles	1.042*** (0.00198)	1.254*** (0.00912)	1.015*** (0.00339)	1.023*** (0.00103)	1.072*** (0.00436)	1.044*** (0.00779)	1.041*** (0.00208)
Observations	384488	223499	137890	745877	141491	46784	198575
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

A prominent finding from our analysis is the significant role of demand-side factors. We find a strong, negative relationship between tweet counts, serving as a proxy for article popularity, and the hazard rate, indicative of an article’s removal from the homepage. Specifically, for the NYT, a 10% increase in tweets is associated with a 0.7% decrease in the hazard rate. This effect is amplified in the WSJ sample, where a 10% increase in tweets corresponds to a 1.4% decrease in hazard rate. This suggests that articles accruing greater attention, as measured by tweet counts, tend to reside on the homepage for longer durations, thereby highlighting the influence of reader demand on the homepage duration of articles.

Turning to supply-side factors, we find their influence to be particularly pertinent for the NYT. In this case, a one-standard-deviation increase in the pro-Democrat score of the article’s abstract, which gauges its political slant, is associated with a 2% decrease in the hazard rate. Interestingly, this association holds across articles of different news tones, suggesting that the editorial decisions of the NYT, as evidenced by article duration on the homepage, are influenced by the article’s political alignment.

In the case of the WSJ, we found the influence of supply-side factors to be negligible. Specifically, the pro-Democrat scores of the articles were not statistically significant in determining their duration on the homepage. This contrasts with the NYT, where political slant played a noticeable role. One potential interpretation of these contrasting findings is that the WSJ might be more motivated by profit-maximizing objectives rather than ideological considerations.

Our results also emphasize the significant role of sentiment in editorial decisions regarding homepage placement. The sentiment of an article, captured through abstract sentiment scores, is notably linked to its duration on the homepage. For the NYT, a one standard deviation increase in the abstract sentiment scores is associated with a 2% decrease in the hazard rate, indicating that more positively toned articles are likely to remain on the homepage longer. The effect is even more pronounced for the WSJ, where a similar increase in sentiment scores corresponds to a 4% decrease in the hazard rate.

Lastly, the publication of other articles by the newspaper within a particular hour is associated with an increased hazard rate. This implies that the introduction of new content can hasten the removal of existing articles from the homepage, reinforcing the dynamic nature of the news cycle. Notice that unlike with a print newspaper front page, there is no hard space constraint on the digital homepage with NYT and WSJ varying the number of articles over time.

6.2 Does Demand Decay Vary by Political Slant?

In our exploration of the influence of supply-side factors on article popularity and longevity, as captured in Table 6, a potential confounder merits attention. Specifically, one might surmise that the observed effects attributed to supply-side factors, as gauged by an article’s pro-Democrat score, might actually be driven by latent demand dynamics. For instance, it could be argued that articles with a pronounced liberal (in the case of NYT) or conservative (in the case of WSJ) slant might inherently elicit greater sustained interest—hence, a more gradual decay in demand—even when prior tweet counts are equivalent.

To rigorously address this concern, we investigate the temporal decay in tweet shares for articles with the following regression:

$$\ln Y_{it} - \ln Y_{it-1} = \beta Z_{it} + \epsilon_{it}, \quad (3)$$

where Y_{it} is the zero-inflated tweets count for article i at time t and therefore $\ln Y_{it} - \ln Y_{it-1}$ is the log difference of tweets count. Z_{it} is a set of (potentially) time-varying controls for article i , including the number of hours since publication, the sentiment and pro-Democrat scores of the article, and number of new articles published by the newspaper at time t . Additionally, the model incorporates several fixed effects dummies for the time of the

article’s publication (hour of the day, day of the week, and year-month).

The estimated constants in Table 7 show that, irrespective of the specific newspaper, tweets manifest a decline ranging from 4.5% to 7.3% per hour during the initial stages after publication. The sole outlier to this pattern is the op-ed articles from NYT. Additionally, as articles age, the rate of decay begins to decelerate, as signified by the estimates on the number of hours since publication.

Table 7: Log Difference (Tweets) Regressions

	NYT				WSJ		
	News	Opinion	Feature	All	News	Feature	All
Sentiment Scores (Abstract)	0.000123 (0.00102)	-0.000479 (0.00290)	-0.000436 (0.00121)	0.000134 (0.000761)	-0.000791 (0.00204)	0.00196 (0.00268)	-0.0000164 (0.00163)
Pro-Dem. Scores (Abstract)	0.0000312 (0.00205)	0.00179 (0.00538)	0.000121 (0.00252)	0.0000425 (0.00153)	0.00236 (0.00406)	0.00502 (0.00541)	0.00313 (0.00326)
Homepage	0.339*** (0.00290)	0.244*** (0.00742)	0.203*** (0.00336)	0.289*** (0.00212)	0.118*** (0.00524)	0.0687*** (0.00680)	0.103*** (0.00418)
Homepage (Last Hour)	-0.353*** (0.00284)	-0.282*** (0.00725)	-0.204*** (0.00337)	-0.297*** (0.00209)	-0.145*** (0.00518)	-0.0904*** (0.00669)	-0.128*** (0.00413)
Hours Published	0.000509*** (0.0000364)	-0.00132*** (0.000101)	0.00182*** (0.0000363)	0.000899*** (0.0000252)	0.000941*** (0.0000763)	0.000808*** (0.0000885)	0.000906*** (0.0000584)
# of New Articles	0.00101*** (0.000162)	0.0137*** (0.00105)	0.00195*** (0.000172)	0.000964*** (0.000117)	0.0000274 (0.000415)	-0.00119** (0.000568)	-0.000794** (0.000369)
Constant	-0.0520*** (0.00367)	0.00619 (0.0102)	-0.0725*** (0.00409)	-0.0446*** (0.00296)	-0.0608*** (0.00655)	-0.0464*** (0.00808)	-0.0603*** (0.00516)
Observations	2214816	324240	1305408	3844464	497136	236352	733488
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Contrary to possible expectations, the pro-Democrat scores reveal no statistically significant association with the rate of tweets decay. In other words, articles that lean more liberally in NYT or more conservatively in WSJ do not exhibit a deceleration in decay distinct from other articles. This observation buttresses our confidence in ascribing the results of Table 6 predominantly to supply-side considerations. Diving deeper into the intrinsic content of the articles, the sentiment scores do not display any marked correlation with the decay in tweets. This suggests that articles, irrespective of whether their tone is positive or negative, follow a consistent demand decay trajectory.

A salient factor influencing tweet decay dynamics is an article’s presence on the homepage. Current placement on the homepage acts as a buffer, attenuating the decay in

demand. Interestingly, having been featured on the homepage during the preceding hour appears to amplify the subsequent decay rate.

Taken together, these findings lend further credence to our primary thesis: that supply-side considerations play a pivotal role in shaping article longevity and popularity on prominent news platforms like NYT and WSJ.

6.3 Survival Analysis by Sections and News Tags

In order to provide more insights into the roles of demand and supply-side factors in shaping the duration of articles on newspaper homepages, we divide our samples by major news section and also by major news tags.

Table 8 shows our results disaggregated by major news sections. Several interesting pattern emerges.

Firstly, the demand-side dynamics, represented by the tweets count, are consistently observed across various news sections. Notably, this effect is stronger in the news sections compared to the feature sections, indicating that the audience engagement, as measured by the number of tweets, exerts a more substantial influence on the survival of articles in the news sections of both newspapers.

Turning to supply-side factors, political slant plays a significant role in the NYT as in Table 6. Liberal-leaning articles tend to have a longer presence on the NYT homepage across several sections, with the notable exception of the style section. This effect is particularly pronounced in the arts section. Contrastingly, for the WSJ, an article’s political slant generally does not bear a statistically significant association with its homepage survival duration. The sole exception to this is the US section, where articles with a pro-Republican slant tend to remain longer on the homepage.

A deeper dive into section-specific dynamics reveals intriguing insights regarding the unique domains of both newspapers. In the case of WSJ, the “C Suite” section, a niche domain particular to the publication, shows hints of supply-side bias. Although the results here are on the brink of statistical significance, they suggest that editorial inclinations might exert a more palpable influence in spaces with reduced competition or specialized content. Conversely, when examining the “NY Region” section of NYT, a distinctive feature of this newspaper, the evidence for any pronounced bias is more tenuous. It is pertinent to note, however, that the sample size for this section was relatively limited,

which could potentially influence the strength of our findings.

Table 8: Survival Analysis of Homepage by Sections (Hazard Ratios)

	NYT						
	Business	US	World	NY Region	Arts	Life	Style
Ln Tweets	0.928*** (0.0125)	0.952*** (0.00621)	0.963*** (0.00894)	0.958 (0.0557)	1.183*** (0.0186)	1.099*** (0.0191)	1.034 (0.0307)
Sentiment Scores (Abstract)	0.948 (0.0349)	1.063*** (0.0198)	0.877*** (0.0209)	0.971 (0.141)	1.031 (0.0336)	1.038 (0.0381)	0.918 (0.0569)
Pro-Dem. Scores (Abstract)	0.979 (0.0682)	0.889*** (0.0337)	0.751*** (0.0418)	1.040 (0.321)	0.654*** (0.0471)	0.838** (0.0683)	1.416*** (0.171)
# of New Articles	1.051*** (0.00559)	1.042*** (0.00311)	1.048*** (0.00417)	1.029 (0.0213)	1.020*** (0.00543)	0.999 (0.00661)	1.018** (0.00882)
Observations	43975	170752	82492	3862	52627	43795	18406

	WSJ						
	Business	US	World	C Suite	Arts	Life	Style
Ln Tweets	0.842*** (0.0195)	0.850*** (0.0171)	0.814*** (0.0156)	1.038 (0.191)	0.937 (0.111)	0.943 (0.0339)	1.210*** (0.0820)
Sentiment Scores (Abstract)	0.989 (0.0518)	0.997 (0.0435)	1.012 (0.0436)	0.781 (0.178)	0.942 (0.102)	0.930 (0.0555)	0.974 (0.127)
Pro-Dem. Scores (Abstract)	1.083 (0.102)	1.282*** (0.118)	0.869 (0.0864)	2.103 (0.981)	0.732 (0.197)	1.004 (0.133)	0.843 (0.223)
# of New Articles	1.066*** (0.00882)	1.085*** (0.00883)	1.074*** (0.00874)	1.073** (0.0370)	1.031 (0.0222)	1.061*** (0.0119)	1.060*** (0.0225)
Observations	33652	33641	32087	23853	4000	20204	5533
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

The role of article sentiment displays an interesting divergence between the two newspapers. In the case of WSJ, the sentiment of an article does not have a significant impact on its duration on the homepage. In contrast, for the NYT, sentiment does play a significant role, particularly in the news section. Generally, articles with a more positive tone are associated with longer survival times on the homepage. However, this trend is inverted in the US news section, where more negative news tend to stay longer.

These results further underscore the complex interplay of demand and supply-side factors in editorial decisions about the duration of an article on a newspaper's homepage. They also highlight the variability of these effects across different types of content and the distinct editorial strategies of the two newspapers studied.

The survival analysis conducted on subsets of articles tagged with keywords corresponding to five major political topics yields further insights into the nature of homepage article

duration, as illustrated in Table 9.¹² The topics encompassed are “Politics”, which includes news about political parties and elections; “Coronavirus”, which covers news on the pandemic and vaccinations; “Race/Gender”, which includes news related to race and gender issues such as discrimination or significant events like the George Floyd incident; “Abortion”, comprising news related to abortion including Supreme Court decisions on the topic; and “Immigration”, which encompasses news pertaining to immigration issues.

Table 9: Survival Analysis of Homepage by Major Tags (Hazard Ratios)

	NYT					WSJ				
	Politics	Coronavirus	RaceGender	Abortion	Immigration	Politics	Coronavirus	RaceGender	Abortion	Immigration
Ln Tweets	0.937*** (0.00665)	1.079*** (0.0135)	1.102*** (0.0145)	0.981 (0.0288)	0.975 (0.0248)	0.971*** (0.00967)	0.878*** (0.0327)	0.905*** (0.0305)	0.749** (0.0894)	0.948 (0.142)
Sentiment Scores (Abstract)	1.012 (0.0210)	1.029 (0.0381)	0.724*** (0.0257)	1.106 (0.104)	0.865** (0.0584)	0.963 (0.0220)	0.889 (0.0974)	0.725*** (0.0582)	1.195 (0.334)	0.856 (0.275)
Pro-Dem. Scores (Abstract)	0.862*** (0.0347)	0.887* (0.0627)	1.107 (0.0877)	0.621** (0.123)	0.577*** (0.0871)	1.216*** (0.0573)	0.997 (0.189)	1.749*** (0.265)	1.139 (0.494)	1.061 (0.695)
# of New Articles	1.023*** (0.00240)	1.023*** (0.00372)	1.017*** (0.00424)	1.015 (0.0102)	1.015* (0.00791)	1.038*** (0.00244)	1.044*** (0.0106)	1.048*** (0.00830)	1.070*** (0.0247)	1.039 (0.0300)
Observations	133830	42765	43969	9245	12121	157417	7888	12823	2299	1170
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Consistent with our prior findings, audience engagement— as indicated by the tweets count— seems to be a key factor associated with a longer duration of articles on the homepage across most topic tags and for both newspapers. The two notable exceptions are for articles with the “Coronavirus” and “Race/Gender” tags in the NYT.

In terms of supply-side factors, a significant relationship was found between political slant and article survival on the homepage. Specifically, in the “Politics” news category, more liberal articles in the NYT and more conservative articles in the WSJ tended to stay longer on the homepage. This relationship is reflected in a 2.4% decrease in hazard rate for a one standard deviation increase in the pro-Democrat score for NYT articles and a 4% increase for WSJ articles. Similar contrasts can be observed across the other topic tags as well, especially “Abortion”.

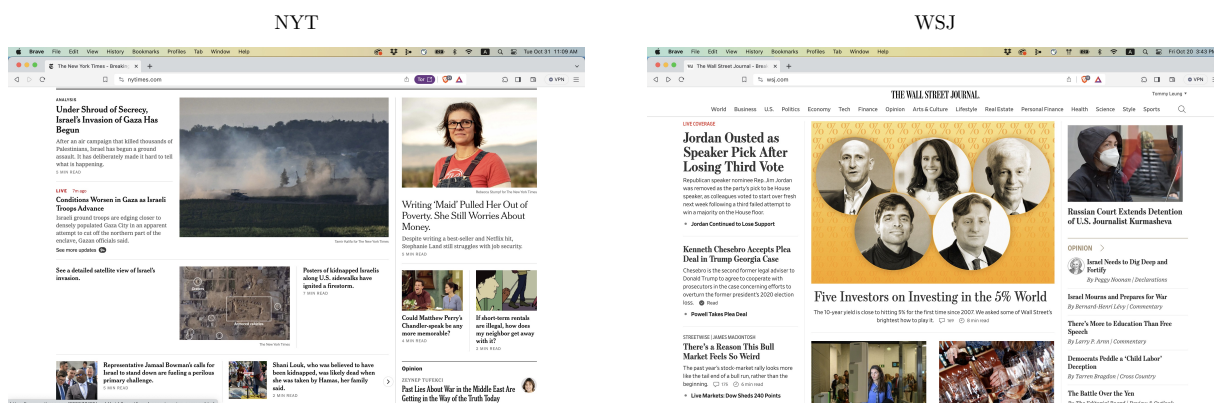
6.4 Survival Analysis on Leading News

While our prior analysis focused broadly on the dynamics of articles across the entirety of the homepages of major newspapers, this section narrows our lens to zero in on a

¹²We conduct a placebo test using several non-politics related news tags in Appendix B.

more exclusive territory: the leading news. These are the articles that occupy the most privileged slots at the top of the homepage, often setting the tone for the day’s narrative and attracting the most immediate attention from visitors. Figure 6 shows position of leading news in NYT and WSJ, respectively. Understanding the survival dynamics of these leading articles offers nuanced insights into the editorial priorities of the publications.

Figure 6: Leading News on NYT and WSJ Homepage



To identify the “leading news” in WSJ, the URL structure of WSJ articles provided a convenient marker for discerning an article’s position on the homepage. Specifically, appended information at the end of each URL demarcates its location. For instance, URLs tagged with “hp_lead” correspond to articles placed among the top 10-11 slots on the homepage—effectively the marquee stories of the hour. Another category, “hp_featst”, denotes articles positioned right below the primary “hp_lead” cluster.¹³ For our analysis, we treated articles marked as “hp_lead” as the leading news for that particular hour in WSJ.

Identifying leading news for the NYT necessitated a distinct approach due to the absence of position-specific markers in their URLs. Instead, the sequence in which URLs were scraped served as our guide. Typically, the first 10 URLs scraped corresponded to their prominent positioning on the NYT homepage. These were thus classified as the leading news for the NYT.

It is important to note that the bulk of these leading articles align with the News category. As such, our examination in this section excludes Opinion and Feature articles, concentrating solely on the News pieces to ensure consistency and specificity in the

¹³The other major position tags in the URLs include “hp_jr”, “hp_lista”, “hp_listb”, “hp_listc”, “hp_minor”, “hp_lead_feature_below_a” and “wsj_hp_buyside”.

findings. The comprehensive results of this refined survival analysis are detailed in Table 10.

Table 10: Survival Analysis of Leading News by Major News Sections

	NYT			WSJ		
	Business	US	World	Business	US	World
Ln Tweets	0.875*** (0.0165)	0.888*** (0.00797)	0.875*** (0.0122)	0.819*** (0.0246)	0.843*** (0.0226)	0.872*** (0.0244)
Sentiment Scores (Abstract)	1.118** (0.0602)	1.158*** (0.0307)	0.987 (0.0356)	0.934 (0.0672)	1.086 (0.0745)	1.074 (0.0757)
Pro-Dem. Scores (Abstract)	1.009 (0.0987)	0.856*** (0.0468)	0.650*** (0.0550)	0.980 (0.132)	0.950 (0.137)	1.610*** (0.268)
# of New Articles	1.054*** (0.00808)	1.031*** (0.00432)	1.032*** (0.00609)	1.096*** (0.0123)	1.074*** (0.0127)	1.058*** (0.0128)
Observations	17901	70081	35486	13909	12347	12037
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Consistent with our results in the previous sections, despite accounting for demand-side dynamics, the influence of political slant remains palpable in the editorial decisions of leading news selection. Within the NYT, liberal-leaning US and World news articles tend to hold their top homepage positions for extended durations. Specifically, an upward shift of one standard deviation in the pro-Democrat scores corresponds to decreases of 3% and 7% in hazard rates for US and World news articles, respectively. This suggests that such articles are less likely to be replaced swiftly, allowing them a more prolonged stay at the pinnacle of the homepage.

On the other hand, the WSJ showcases a distinct trend for its World news section. Articles with a conservative tilt are more resilient in their top positions. Here, a one standard deviation rise in the pro-Democrat scores is linked with a 12% increase in hazard rates, indicating that liberal-leaning World news articles in the WSJ are more susceptible to quicker replacement from the prime spots.

7 Editorial Decisions on Print Edition and Podcasts

Understanding editorial decisions in news media has grown increasingly important as these choices reflect both journalistic values and market considerations. Gentzkow and Shapiro (2006) argue that competition can reduce media bias and improve reporting accuracy. A

central reason is the presence of numerous media outlets, which ensures that readers often have access to diverse follow-up stories, making it easier to verify initial reports.

This notion of competition’s influence on media bias is further explored by Gentzkow, Shapiro, and Stone (2015) and Puglisi and Snyder (2015). Their reviews of the literature suggest that competition generally reduces distortions, aligning media outcomes more closely with a bias-free state.

Drawing from these insights, a key observation emerges: media bias might be influenced by the elasticity of demand. Media facing inelastic demand might exhibit greater bias, while those facing elastic demand might lean towards reduced bias. This distinction underpins our examination of editorial decisions at the print edition of the NYT and podcasts like “The Daily” (NYT) and “The Journal” (WSJ). The print edition, typically consumed by a consistent readership, is likely to face inelastic demand. In contrast, podcasts, aiming to capture new audiences, operate in an environment with more elastic demand. By analyzing editorial choices across these two platforms, we seek to understand how demand elasticity and competition influence content decisions.

7.1 Print Edition

In our examination of editorial decisions regarding articles in the print edition, we primarily focus on the NYT due to data availability constraints. A salient indicator we used from the API is the presence of a “print headline” for an article, suggesting its appearance in the print edition. Unfortunately, this metric was unavailable for our WSJ dataset, hence restricting our analysis to the NYT sample.

We manually checked the headline from the API against the one listed at the bottom of articles on the NYT website. The “print headline” variable serves as an accurate representation of an article’s inclusion in the print edition up to February 2023. Historically, between 60% to 70% of articles (per month) were showcased in the print edition, but starting February 2023, this figure fell to below 40%. Given this data anomaly, we limited our analysis to include only articles up to January 2023, leaving a sample size of 63,740 articles.

Our preliminary results, presented in Table 11, consider whether an article is included in the print edition. We use a linear regression model where the dependent variable is the binary outcome if an article appears in the print edition. Articles that have been featured

on the homepage at some point, or those that have garnered a significant number of tweets, exhibit a higher probability of being selected for the print edition. Interestingly, the article’s pro-Democrat scores do not show a significant correlation with its inclusion in the print version.

Table 11: Print Edition Linear Regression by News Tone

	(1) News	(2) Opinion	(3) Feature	(4) All
Ever on Homepage	0.364*** (0.00534)	0.190*** (0.0322)	0.0845*** (0.00691)	0.251*** (0.00421)
In Tweets Count	0.0556*** (0.00158)	0.0764*** (0.00489)	0.0832*** (0.00289)	0.0685*** (0.00134)
Sentiment Scores (Abstract)	0.0442*** (0.00489)	-0.0385** (0.0154)	0.00304 (0.00757)	0.0186*** (0.00402)
Pro-Dem. Scores (Abstract)	0.00499 (0.00972)	0.0224 (0.0284)	-0.0184 (0.0156)	0.00439 (0.00802)
Observations	37560	5508	20672	63740
Day FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Tone FE	No	No	No	Yes

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

A more granular exploration of specific news sections yields more nuanced insights, as depicted in Table 12. It appears that supply-side factors come to the fore in prominent news sections such as “US”, “Business”, and in the features sections “Arts”, and “Style”. For instance, in the “Arts” section, an article’s pro-Democrat score plays a notable role: a one standard-deviation surge in this score correlates with a 3.4% increase in the probability of its appearance in the print edition, even after accounting for the article’s popularity based on tweet count. Comparable patterns, albeit with slightly lesser magnitudes ranging from nearly 1% to 2%, are observable in the “Style”, “Business”, and “US” sections. These findings underscore the multifaceted nature of editorial decisions, where both demand and supply-side dynamics shape the selection of articles for the print edition.

Table 12: Print Edition Linear Regression by Major NYT News Sections

	(1)	(2)	(3)	(4)	(5)	(6)
	Business	US	World	Arts	Life	Style
Ever on Homepage	0.277*** (0.0137)	0.461*** (0.00977)	0.540*** (0.0104)	-0.0406*** (0.00875)	0.152*** (0.0141)	0.0322* (0.0183)
ln Tweets	0.0656*** (0.00498)	0.0659*** (0.00265)	0.0602*** (0.00391)	0.0232*** (0.00437)	0.0163*** (0.00581)	0.0422*** (0.00838)
Sentiment Scores (Abstract)	0.0452*** (0.0138)	0.0390*** (0.00798)	-0.0287*** (0.00906)	-0.0134 (0.00936)	0.0141 (0.0147)	-0.0286 (0.0219)
Pro-Dem. Scores (Abstract)	0.0571** (0.0257)	0.0350** (0.0154)	-0.0121 (0.0208)	0.170*** (0.0208)	-0.0304 (0.0309)	0.101** (0.0422)
Observations	4967	12252	7761	8736	4874	2531
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

7.2 Podcasts

The rapidly increasing popularity of podcasts as a medium for news consumption cannot be understated. This transition is accentuated by Meredith Kopit Levien, the Chief Executive of The New York Times, who noted the importance of The Daily, the newspaper’s flagship podcast, when presenting the Times’ third-quarter results in 2020. She cites its audience of 4 million daily, a figure that is “almost twice as large as the paper was at its peak” (Turvill, 2020).

Moreover, Levien underscores the difference in the demographic profile of this newfound audience. In the press conference, she highlighted that the majority of The Daily’s listeners are under the age of 50, with a substantial number even younger, below 40. This demographic shift inevitably diversifies the range of news that resonates with this age group. Levien’s observation that this younger cohort has developed “an affinity for The New York Times” signifies the podcast’s role in transforming news consumption patterns. Notice this younger readers disproportionately use news aggregators such as Google News or Facebook News Feed. For instance, the Reuters Institute Digital News Report in 2023 finds that “younger users are less likely to go directly to a news site or app and more likely to use social media or other intermediaries.”¹⁴ These sites source news for a wide

¹⁴See <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/dnr-executive-summary>.

range of sites, and so are potentially a causal factor in reducing loyalty to any given news source.

Beyond attracting a large listenership, podcasts have also begun playing a central role in solidifying brand loyalty and amplifying newspaper subscriptions. Although Levien acknowledges the challenges in directly attributing subscription growth to the podcast, she remains optimistic. She remarked on the podcast’s instrumental role, saying, “It’s helpful in driving affinity to the brand. Given the improved results since the inception of The Daily four years ago, there’s every reason to believe in its impact.”

Our research took this evolving landscape into account and delved deep into understanding the editorial decisions that drive article selection for flagship podcasts. Specifically, our analysis focused on articles featured in The Daily by the NYT and The Journal by the WSJ during our sample period. To acquire this data, we visited the URLs of the respective NYT and WSJ podcasts daily, and extracted the URLs of the “feature articles” presented within each podcast episode. These articles were then matched back to our primary articles dataset. It is crucial to note that only those articles featured in a podcast within 30 days of their original publication were considered for our analysis.¹⁵

Building on this data, our findings, detailed in Table 13, underscore distinct patterns that emerged from employing a linear regression model akin to our print edition evaluation. A salient observation from this analysis is the absence of any op-ed articles from the NYT being spotlighted in The Daily. As a result, the column dedicated to Opinion articles was omitted from our dataset.¹⁶

¹⁵This constitutes more than 90% of articles that featured in the podcasts in our sample.

¹⁶We conduct a similar analysis by news sections in Appendix D.

Table 13: Podcast Linear Regression by News Tones

	NYT			WSJ		
	News	Feature	All	News	Feature	All
Ever on Homepage	0.000801 (0.00137)	0.000355 (0.000561)	0.000741 (0.000818)	0.000138 (0.00413)	-0.00305 (0.00334)	-0.00275 (0.00238)
In Tweets	0.00670*** (0.000403)	0.00165*** (0.000234)	0.00510*** (0.000260)	0.0143*** (0.00108)	0.0116*** (0.00109)	0.00926*** (0.000628)
Sentiment Scores (Abstract)	-0.00353*** (0.00124)	-0.000792 (0.000614)	-0.00275*** (0.000776)	-0.00454 (0.00303)	-0.000527 (0.00333)	-0.00307 (0.00201)
Pro-Dem. Scores (Abstract)	-0.000248 (0.00247)	-0.00126 (0.00127)	-0.000293 (0.00155)	0.00383 (0.00603)	-0.0109 (0.00679)	-0.000277 (0.00392)
Observations	41241	22957	70287	10289	4630	17755
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Tone FE	No	No	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

While the left panel of the table presents results for the NYT and the right does so for the WSJ, a consistent trend emerges across both newspapers: the count of tweets an article receives, indicative of its popularity, tends to correlate positively with its chance of being featured in a podcast. Interestingly, the supply-side factors, exemplified by pro-Democrat scores, do not seem to significantly influence the podcast feature decisions. This observation is consistent with the hypothesis suggesting a reduced supply-side media bias in settings where demand is more elastic. This trend remains consistent even when data is examined by major news sections instead of by news tone.

Additionally, there seems to be no significant connection between articles being displayed on the homepage and their selection for podcast features. However, when it comes to sentiment, articles that adopt a more negative tone seem to find favor with The Daily, the flagship podcast of the NYT.

8 Conclusion

In synthesizing our exploration into the dynamics of media bias within the digital front pages of two renowned national newspapers, we arrive at a series of salient insights and corresponding limitations. Drawing from an expansive dataset of articles from both The New York Times (NYT) and The Wall Street Journal (WSJ), juxtaposed against a backdrop of Twitter interactions, we are presented with a nuanced understanding of the underpinnings influencing contemporary editorial decisions.

Central to our observations is the dynamic interplay between audience engagement (the demand side) and the innate political orientations of articles (the supply side). The platform of a homepage not only symbolizes an article's prominence but actively contributes to its reach, as manifested in heightened tweet engagements.

Nevertheless, this symbiotic relationship between content and its audience extends beyond mere engagement metrics. Our machine learning-driven approach unearths latent biases affecting the homepage longevity of articles. The propensity of the NYT to favor liberal-leaning articles, contrasted against the WSJ's predilection for conservative content, unravels the intricate balance newspapers strike between satisfying readership tastes and adhering to their editorial proclivities.

Broadening our scope, a delve into print and podcast mediums brought to light another pivotal observation: media outlets, when confronted with a more versatile audience demand, seem to dial back their inherent biases, arguably in a bid to capture a wider readership spectrum.

However, as we close this chapter of inquiry, it is imperative to acknowledge inherent caveats. Firstly, our lens is trained on two contrasting national newspapers, leaving out the rich tapestry of local media that many in the scholarly community have delved into. Secondly, while we harness the internal political slant variations within a newspaper to unearth supply-side biases, we are making a presumption of their exogeneity. True, every newspaper carries its editorial compass, as echoed in the overarching political stances of our chosen publications. Yet, we argue that the randomness of news events imparts a certain exogenous quality to an article's tone, especially when gauged against the newspaper's broader ideological milieu.

In summation, our study, with its deep dives and recognized limitations, shines a light on the multifaceted realm of media bias and digital strategy interplay. It also underscores the need for further inquiry as the contours of media consumption continue to evolve, ensuring the pursuit of transparency and fairness within our primary information channels.

References

- Aguiar, Luis, Joel Waldfogel, and Sarah Waldfogel. 2021. “Playlisting favorites: Measuring platform bias in the music industry.” *International Journal of Industrial Organization* 78:102765.
- Allcott, Hunt and Matthew Gentzkow. 2017. “Social Media and Fake News in the 2016 Election.” *Journal of Economic Perspectives* 31 (2):211–236.
- Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. 2020. “Evaluating the fake news problem at the scale of the information ecosystem.” *Science Advances* 6 (14):eaay3539.
- Anderson, Monica, Lee Rainie, and Haley Nolan. 2021. “Social Media Use in 2021.” Tech. rep., Pew Research Center.
- Anderson, Simon P. and Joshua S. Gans. 2011. “Platform Siphoning: Ad-Avoidance and Media Content.” *American Economic Journal: Microeconomics* 3 (4):1–34.
- Ansolabehere, Stephen, Rebecca Lessem, and James M. Snyder. 2006. “The Orientation of Newspaper Endorsements in U.S. Elections, 1940–2002.” *Quarterly Journal of Political Science* 1 (4):393–404.
- Armstrong, Mark and Jidong Zhou. 2011. “PAYING FOR PROMINENCE.” *The Economic Journal* 121 (556):F368–F395.
- Athey, Susan and Michael Luca. 2019. “Economists (and Economics) in Tech Companies.” *Journal of Economic Perspectives* 33 (1):209–230.
- Baron, David P. 2006. “Persistent media bias.” *Journal of Public Economics* 90 (1):1–36.
- Baum, Matthew A. and Tim Groeling. 2008. “New Media and the Polarization of American Political Discourse.” *Political Communication* 25 (4):345–365.
- Bennet, James. 2023. “When the New York Times lost its way.” *The Economist* URL <https://www.economist.com/1843/2023/12/14/when-the-new-york-times-lost-its-way>.

- Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.
- Bourreau, Marc and Germain Gaudin. 2022. "Streaming platform and strategic recommendation bias." *Journal of Economics & Management Strategy* 31 (1):25–47.
- Boxell, Levi and Jacob Conway. 2022. "Journalist Ideology and the Production of News: Evidence from Movers." Working Paper.
- Brenan, Megan. 2022. "Americans' Trust In Media Remains Near Record Low." *Gallup* URL <https://news.gallup.com/poll/403166/americans-trust-media-remains-near-record-low.aspx>.
- Cagé, Julia, Moritz Hengel, Nicolas Hervé, and Camille Urvoy. 2022. "Hosting Media Bias: Evidence from the Universe of French Broadcasts, 2002-2020." URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4036211.
- Cagé, Julia, Nicolas Hervé, and Béatrice Mazoyer. 2022. "Social Media Influence Mainstream Media: Evidence from Two Billion Tweets." URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3663899. Working Paper.
- Calzada, Joan, Nestor Duch-Brown, and Ricard Gil. 2022. "Do Search Engines Increase Concentration in Media Markets?" Working paper.
- Calzada, Joan and Ricard Gil. 2020. "What Do News Aggregators Do? Evidence from Google News in Spain and Germany." *Marketing Science* 39 (1):134–167.
- Chan, Jimmy and Wing Suen. 2008. "A Spatial Theory of News Consumption and Electoral Competition." *The Review of Economic Studies* 75 (3):699–728.
- Chen, Nan and Hsin-Tien Tiffany Tsai. 2023. "Steering via Algorithmic Recommendations." URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3500407.
- Claussen, Jorg, Christian Peukert, and Ananya Sen. 2023. "The Editor and the Algorithm: Returns to Data and Externalities in Online News." *Management Science* Forthcoming.
- de Cornière, Alexandre and Miklos Sarvary. 2022. "Social Media and News: Content Bundling and News Quality." *Management Science* :forthcoming.

- de Cornière, Alexandre and Greg Taylor. 2019. “A model of biased intermediation.” *The RAND Journal of Economics* 50 (4):854–882.
- Demsetz, Harold and Kenneth Lehn. 1985. “The Structure of Corporate Ownership: Causes and Consequences.” *Journal of Political Economy* 93 (6):1155–1177.
- Dias, Elizabeth and Jack Healy. 2022. “For Many Who Marched, Jan. 6 Was Only the Beginning.” *The New York Times* URL <https://www.nytimes.com/2022/01/23/us/jan-6-attendees.html>.
- Dobbins, James, J. David Goodman, and Edgar Sandoval. 2022. “At Least 46 Migrants Found Dead in Tractor-Trailer in San Antonio.” *The New York Times* URL <https://www.nytimes.com/2022/06/28/us/migrants-san-antonio-tractor-killed.html>.
- Farronato, Chiara, Andrey Fradkin, and Alexander MacKay. 2023. “Self-Preferencing at Amazon: Evidence from Search Rankings.” *AEA Papers and Proceedings* 113:239–43. URL <https://www.aeaweb.org/articles?id=10.1257/pandp.20231068>.
- Gentzkow, Matthew, Edward L. Glaeser, and Claudia Goldin. 2006. *The Rise of the Fourth Estate. How Newspapers Became Informative and Why It Mattered*. University of Chicago Press, 187–230. URL <https://www.nber.org/books-and-chapters/corruption-and-reform-lessons-americas-economic-history>.
- Gentzkow, Matthew and Jesse M. Shapiro. 2008. “Competition and Truth in the Market for News.” *Journal of Economic Perspectives* 22 (2):133–154.
- . 2010. “What Drives Media Slant? Evidence from U.S. Daily Newspaper.” *Econometrica* 78 (1):35–71.
- Gentzkow, Matthew, Jesse M. Shapiro, and Daniel F. Stone. 2015. “Chapter 14 - Media Bias in the Marketplace: Theory.” In *Handbook of Media Economics, Handbook of Media Economics*, vol. 1, edited by Simon P. Anderson, Joel Waldfogel, and David Strömberg. North-Holland, 623–645.
- Gentzkow, Matthew and Jesse M. Shapiro. 2006. “Media Bias and Reputation.” *Journal of Political Economy* 114 (2):280–316.

- George, Lisa M. and Joel Waldfogel. 2006. “The New York Times and the Market for Local Newspapers.” *American Economic Review* 96 (1):435–447.
- Groseclose, Tim and Jeffrey Milyo. 2005. “A Measure of Media Bias*.” *The Quarterly Journal of Economics* 120 (4):1191–1237.
- Hagar, Nick and Nicholas Diakopoulos. 2019. “Optimizing Content with A/B Headline Testing: Changing Newsroom Practices.” *Media and Communication* 7 (1):117–127.
- Hagiu, Andrei and Bruno Jullien. 2014. “Search diversion and platform competition.” *International Journal of Industrial Organization* 33:48–60.
- Hatte, Sophie, Etienne Madinier, and Ekaterina Zhuravskaya. 2022. “Reading Twitter in the Newsroom: Web 2.0 and Traditional-Media Reporting of Conflicts.” URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3845739.
- Hirano, Shigeo and James M. Snyder. 2020. “Measuring the Partisan Behavior of U.S. Newspapers, 1880 to 1980.”
- Huang, Yihong, Juanjuan Meng, and Xi Weng. 2022. “Strategic Thinking and Media Bias: Evidence from Chinese Microblog Users.” URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4184003. Working Paper.
- Hutto, Clayton and Eric Gilbert. 2014. “Vader: A parsimonious rule-based model for sentiment analysis of social media text.” *Proceedings of the international AAAI conference on web and social media* 8 (1):216–225.
- Jeon, Doh-Shin and Nikrooz Nasr. 2016. “News Aggregators and Competition among Newspapers on the Internet.” *American Economic Journal: Microeconomics* 8 (4):91–114.
- Jürgensmeier, Lukas and Bernd Skiera. 2023. “Measuring Fair Competition on Digital Platforms.”
- Leung, Tin Cheuk and Koleman Strumpf. 2023. “All the Headlines that Are Fit to Change.” URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4075398. Working Paper.

- Martin, Gregory J. and Joshua McCrain. 2019. “Local News and National Politics.” *American Political Science Review* 113 (2):372–384.
- Martin, Gregory J. and Ali Yurukoglu. 2017. “Bias in Cable News: Persuasion and Polarization.” *American Economic Review* 107 (9):2565–99.
- Muise, Daniel, Homa Hosseinmardi, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. 2022. “Quantifying partisan news diets in Web and TV audiences.” *Science Advances* 8 (28):eabn0083.
- Mullainathan, Sendhil and Andrei Shleifer. 2005. “The Market for News.” *American Economic Review* 95 (4):1031–1053.
- New York Times. 2022. “The New York Times Company 2022 Annual Report.” Tech. rep., The New York Times Company. URL <https://nytco-assets.nytimes.com/2023/03/The-New-York-Times-Company-2022-Annual-Report.pdf>.
- Olmstead, Kenny, Amy Mitchell, and Tom Rosenstiel. 2011. “Navigating News Online: Where People Go, How They Get There and What Lures Them Away.” Tech. rep., Pew Research Center. URL <https://www.pewresearch.org/wp-content/uploads/sites/8/legacy/NIELSEN-STUDY-Copy.pdf>.
- Parker, Geoffrey, Georgios Petropoulos, and Marshall W. Van Alstyne. 2020. “Digital Platforms and Antitrust.” URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3608397.
- Puglisi, Riccardo and James M. Snyder. 2011. “Newspaper Coverage of Political Scandals.” *The Journal of Politics* 73 (3):931–950.
- . 2015. “Chapter 15 - Empirical Studies of Media Bias.” In *Handbook of Media Economics, Handbook of Media Economics*, vol. 1, edited by Simon P. Anderson, Joel Waldfogel, and David Strömberg. North-Holland, 647–667.
- Qin, Bei, David Strömberg, and Yanhui Wu. 2018. “Media Bias in China.” *American Economic Review* 108 (9):2442–2476.
- Raval, Devesh. 2023. “Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box.” URL <https://deveshraval.github.io/buyBox.pdf>.

Reimers, Imke and Joel Waldfogel. 2023. “A Framework for Detection, Measurement and Welfare Analysis of Platform Bias.” Working Paper.

Shapiro, Jesse M. 2016. “Special interests and the media: Theory and an application to climate change.” *Journal of Public Economics* 144:91–108.

Shearer, Elisha. 2020. “Two-thirds of U.S. adults say they’ve seen their own news sources report facts meant to favor one side.” *Pew Research Center* URL <https://www.pewresearch.org/short-reads/2020/11/02/two-thirds-of-u-s-adults-say-theyve-seen-their-own-news-sources-report-facts-meant-to-favor-one-side/>

Strömberg, David. 2004. “Mass Media Competition, Political Competition, and Public Policy.” *The Review of Economic Studies* 71 (1):265–284. URL <http://www.jstor.org/stable/3700719>.

Turvill, William. 2020. “How The Daily podcast is helping The New York Times drive advertising and subscription growth.” *Press Gazette* URL <https://pressgazette.co.uk/news/how-the-daily-podcast-is-helping-the-new-york-times-drive-advertising-and-subscription-growth/>

Appendix A WordClouds of Other News Sections

An examination of word clouds for news articles within the ‘World’ section in Figure A1 reveals distinct patterns related to geopolitical events and narratives. A salient observation is the frequent mention of the Ukraine conflict across both newspapers. Notably, while this theme appears in both liberal and conservative articles, there’s a marked preponderance in liberal articles in both newspapers. This might indicate a broader focus or perhaps a specific narrative angle that liberal articles adopt when discussing the Ukraine situation.

Figure A1: Word Clouds of World News Articles’ Abstracts



Furthermore, in relation to China, there’s a discernible pattern. Keywords such as “China”, “Xi”, and “Communist” manifest with greater frequency in conservative articles across both newspapers. This prevalence aligns with the hawkish stance against China that has been notable among Republican politicians.

In analyzing the word clouds for articles in the ‘Business’ section, Figure A2 specific thematic trends become evident, elucidating the ideological bent of the content. Firstly,

Appendix B Survival Analysis by News Tags (Placebo)

The results from our survival analysis documented in Table 9 provides insights from subsets of articles tagged with keywords corresponding to five major political topics. A clear takeaway is the importance of supply-sided factors in decoding media bias. In contrast to this primary observation, this section endeavors to execute a placebo test, targeting articles distinguished by non-political tags.

To establish the relative insignificance of political bias in non-political domains, we closely examined articles categorized under tags such as “Cookings,” which encompassed keywords like “Cookbooks”, “Recipes”, and “Restaurants”; “Fashion” that included the terms “Fashion” and “Style”; and “Sports” which amalgamated all sports-related tags.

Given their inherent nature, articles under these non-political tags would be anticipated to manifest reduced susceptibility to supply-side biases, especially when juxtaposed against the politically tinted articles from Table 9. Indeed, Table A1 shows that, the pro-Democrat scores, symbolic of an article’s political leaning, do not show a statistically significant correlation with the duration of articles on either newspaper’s homepage for these non-political categories. However, it is worth noting an intriguing outlier: liberal articles tagged under “Cookings” in the WSJ tend to exhibit a trend of prolonged homepage presence compared to our principal findings, though this observation remains statistically insignificant.

Table A1: Survival Analysis of Homepage by Other Tags (Hazard Ratios)

	NYT			WSJ		
	Cookings	Fashion	Sports	Cookings	Fashion	Sports
Ln Tweets	0.890*** (0.0370)	1.081* (0.0492)	0.928** (0.0274)	0.874 (0.0770)	0.911*** (0.0322)	1.007 (0.0425)
Sentiment Scores (Abstract)	0.867 (0.0762)	0.796* (0.0939)	0.729*** (0.0420)	1.155 (0.205)	0.974 (0.0642)	0.968 (0.0668)
Pro-Dem. Scores (Abstract)	0.962 (0.163)	1.056 (0.225)	1.171 (0.158)	0.619 (0.181)	0.944 (0.132)	1.114 (0.170)
# of New Articles	1.008 (0.00826)	1.016 (0.0103)	1.001 (0.00689)	1.041*** (0.0154)	1.037*** (0.00632)	1.034*** (0.00704)
Observations	9117	7343	16316	4798	23675	13938
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Appendix C Survival Analysis of Leading News by Major News Tags

In our primary analysis, presented in Section 6.4 and depicted in Table 10, we uncovered evidence of supply-side bias for leading news within major news sections, especially for categories such as World and US news prominently displayed on the homepages of both newspapers. To fortify these findings, this section endeavors to offer an additional layer of robustness checks, probing the patterns associated with specific major news tags.

Delving into the details, Table A2 catalogues these outcomes. The results exhibit that on the NYT homepage, articles labeled with tags such as “Politics”, “Coronavirus”, “Abortion”, and “Immigration” that have a liberal bent tend to dominate the top spots for prolonged periods.

Table A2: Survival Analysis of Leading News by Major Tags (Hazard Ratios)

	NYT					WSJ				
	Politics	Coronavirus	Race/Gender	Abortion	Immigration	Politics	Coronavirus	Race/Gender	Abortion	Immigration
Ln Tweets	0.937*** (0.00665)	1.079*** (0.0135)	1.102*** (0.0145)	0.981 (0.0288)	0.975 (0.0248)	0.971*** (0.00967)	0.878*** (0.0327)	0.905*** (0.0305)	0.749** (0.0894)	0.948 (0.142)
Sentiment Scores (Abstract)	1.012 (0.0210)	1.029 (0.0381)	0.724*** (0.0257)	1.106 (0.104)	0.865** (0.0584)	0.963 (0.0220)	0.889 (0.0974)	0.725*** (0.0582)	1.195 (0.334)	0.856 (0.275)
Pro-Dem. Scores (Abstract)	0.862*** (0.0347)	0.887* (0.0627)	1.107 (0.0877)	0.621** (0.123)	0.577*** (0.0871)	1.216*** (0.0573)	0.997 (0.189)	1.749*** (0.265)	1.139 (0.494)	1.061 (0.695)
# of New Articles	1.023*** (0.00240)	1.023*** (0.00372)	1.017*** (0.00424)	1.015 (0.0102)	1.015* (0.00791)	1.038*** (0.00244)	1.044*** (0.0106)	1.048*** (0.00830)	1.070*** (0.0247)	1.039 (0.0300)
Observations	133830	42765	43969	9245	12121	157417	7888	12823	2299	1170
Hour FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Exponentiated coefficients; Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Conversely, for the WSJ, a different pattern surfaces. Leading news articles tagged with “Politics” and “Race/Gender”, and that exude a more conservative orientation, demonstrate resilience in retaining their paramount positions on the homepage.

Appendix D Podcast Analysis by News Sections

In our endeavor to comprehend the nuanced interplay between media bias and editorial choices, we previously illustrated in Table 13 a seemingly weak linkage between an article’s political slant and its selection for inclusion in a podcast. Deepening this analysis, we now examine articles classified by their respective news sections rather than by their news tone. The outcomes of this refined exploration are encapsulated in Table A3.¹⁷

Table A3: Podcast Linear Regression by Major News Sections

	NYT					
	Business	US	World	Arts	Life	Style
Ever on Homepage	0.00205 (0.00279)	-0.00120 (0.00322)	0.00760** (0.00354)	0.000434 (0.000730)	0.000636 (0.00174)	0 (.)
In Tweets	0.00429*** (0.00101)	0.00865*** (0.000858)	0.00852*** (0.00132)	0.00168*** (0.000363)	0.00291*** (0.000722)	0 (.)
Sentiment Scores (Abstract)	0.00594** (0.00278)	0.000923 (0.00258)	-0.00925*** (0.00303)	-0.000621 (0.000778)	-0.00232 (0.00182)	0 (.)
Pro-Dem. Scores (Abstract)	-0.00834 (0.00518)	0.00149 (0.00499)	-0.00321 (0.00701)	0.000446 (0.00174)	-0.00766** (0.00383)	0 (.)
Observations	5493	13441	8585	9735	5380	2830

	WSJ					
	Business	US	World	Arts	Life	Style
Ever on Homepage	-0.0141 (0.00992)	0.00424 (0.00785)	-0.00103 (0.00701)	0 (.)	0.000135 (0.00395)	0.00266 (0.00811)
In Tweets	0.0190*** (0.00257)	0.00747*** (0.00190)	0.0119*** (0.00187)	0 (.)	0.00386*** (0.00130)	0.00176 (0.00270)
Sentiment Scores (Abstract)	-0.0190** (0.00828)	-0.000467 (0.00542)	-0.00255 (0.00522)	0 (.)	0.00510 (0.00333)	-0.00590 (0.00760)
Pro-Dem. Scores (Abstract)	0.0118 (0.0146)	-0.00133 (0.0103)	-0.0118 (0.0113)	0 (.)	-0.0000768 (0.00696)	0.0257* (0.0151)
Observations	2589	2509	2614	1042	1599	507
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

In evaluating the broader dataset, a consistent theme emerges which reaffirms our initial findings from Table 13: the political bias of an article does not appear to significantly influence its candidacy for podcast features. However, embedded within this overarching narrative are intriguing exceptions. Specifically, the “Life” section articles from the NYT and the “Style” segment articles from the WSJ deviate from this pattern. In a somewhat

¹⁷Certain sections, namely “Style” for the NYT and “Arts” for the WSJ, are entirely absent from their corresponding podcast features.

counterintuitive twist, it is the conservative articles from NYT's "Life" section and the liberal articles from WSJ's "Style" section that are more likely to feature in their respective podcasts. This unanticipated trend, contrary to the conventional political inclinations of the newspapers, provides credence to the hypothesis that newspapers may be utilizing podcasts to resonate with a more diverse listenership, potentially reaching outside their traditional political constituencies.