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

Tin Cheuk Leung† Wake Forest University

Koleman Strumpf‡ Wake Forest University

September 2024

Abstract

In this study we propose a novel approach to detect supply-side media bias, which does not require information about the ideological leanings of owners or editors. We focus on the most visible and read part of a media site, its digital homepage. Analyzing over 100,000 articles from The New York Times (NYT) and The Wall Street Journal (WSJ), complemented by data from 24 million tweets, we assess the factors influencing article duration on their homepages. After flexibly controlling for demand-side preferences, we attribute extended homepage presence of ideologically slanted articles to supply-side biases. Both demand (tweets count) and supply (a machine learning measure of article ideological orientation) significantly impact homepage longevity. Supply-side bias leads liberal articles 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, supporting the presence of supply-side biases.

JEL Classification: D22, D72, D83, L82
Keywords: Media Bias, Media Economics, Social Media, Machine Learning

*We would like to thank Julia Cage, Darlene C. Chisholm, Shane Greenstein, Tim Groseclose, David Rothschild, Jose De Sousa, seminar participants from APIOS, CESifo Conference on Economics of Digitization, CREED, CUHK, Digital Economy Workshop Rotterdam, HKEA Biennial Conference, IIOC, The Econometric Society NASM, Paris Conference on Digital Economics, SEA for their comments, Sarah Bahr for information about the New York Times print deadline, 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.

†Contact: Department of Economics, Kirby Hall, Wake Forest University, Box 7505, Winston-Salem, NC 27109. Email: leungtc@wfu.edu

‡Email: strumpks@wfu.edu

1 Introduction

Media bias is a prevalent issue globally, supported by extensive research and corresponding low public trust in the media.[1] There are also detailed reports from within major media sources. Bennet (2023) critically examines bias at the New York Times (NYT).[2] As a former NYT op-ed editor, Bennet (2023) discusses potential causes for the shift in bias at the NYT, attributing it to both a demand-driven shift from advertising to subscription revenue models and a supply-driven change in reporters' ideals, reducing their focus on objectivity. Understanding these bias sources is crucial given the ambiguous effects competition may exert. The theoretical framework outlined by Gentzkow, Shapiro, and Stone (2015) suggests, "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 means supply- or demand-driven bias will have different implications for antitrust regulators.

This paper introduces a novel method for decomposing media bias that concentrates on supply-side factors independent of a newspaper's ownership or editorial biases.[3] Our approach evaluates the factors which shape article duration on a newspaper's digital homepage. Homepage articles have a privileged position, as it is the starting point for most readers which in turn drives views and shares of the featured articles. One of the most important editorial decisions is selecting how long each article remains at this prized location. After adjusting for demand-side effects using reader engagement data, we use prolonged homepage presence of ideologically slanted articles as indicative of supply-side biases. This method is versatile and can be applied to any media site using only data on homepage duration, story content, and commonly available reader analytics.

To motivate our analysis, consider the coverage by The NYT and The Wall Street Journal (WSJ) on November 16, 2022, regarding 12 Republicans supporting same-sex

[1] Brenan (2022) reports that only 34% of Americans believe the mass media reports news "fully, accurately, and fairly," with just 7% expressing substantial trust and confidence, and 27% a moderate amount. Additionally, Pew Research Center (2020) notes that two-thirds of U.S. adults perceive a bias in news reporting, often favoring a particular political orientation. Pew Research Center (2018) observes that "trust in the news media dips lowest in Spain, France, the UK, and Italy, with roughly a quarter of people with populist views in each country expressing confidence in the news media."

[2] Folkenflik (2020, 2024) provide similar insider accounts of the decline in viewpoint diversity at the Los Angeles Times and NPR.

[3] In 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."

marriage legislation in the Senate. Both articles displayed a liberal-leaning perspective. The NYT portrayed the event as a sign of shifting attitudes within the Republican Party, emphasizing internal divisions (Astor, 2022). The WSJ, generally seen as conservative, underscored the bipartisan effort, remarking on the collaboration across party lines with, "A dozen Republicans join Democrats in vote, which opens effort to pass bill in both houses of Congress by end of year." (Ferek, 2022). Both articles were published simultaneously at around 6pm. Despite similar reader interest and liberal tone, the duration on their respective homepages varied significantly. The NYT article stayed on its homepage for 12 hours, while the WSJ article was removed after just 4 hours. This discrepancy, given their comparable content and engagement, suggests a potential role of supply-side media bias by the editors of the newspapers.

This study utilizes a unique dataset comprising articles featured on the homepages of the two largest daily newspapers in the United States, The NYT and WSJ, as identified by recent statistics (Statista, 2024).[4] The dataset spans from August 2021 to May 2023 for the NYT and from October 2022 to May 2023 for the WSJ. The dataset includes over 80,000 articles from the NYT and more than 20,000 from the WSJ, augmented by Twitter data collected via the Twitter API as a measure of reader engagement. Specifically, our analysis covers approximately 22 million tweets linked to NYT articles (averaging about 275 tweets per article) and nearly 2 million tweets for WSJ articles (averaging about 100 tweets per article).

To analyze the sentiment and political orientations of the articles, we used textual analysis to calculate sentiment scores and pro-Democrat scores. These pro-Democrat scores, which reflect an article's political slant, were derived using a machine-learning model trained on tweets from approximately 2,000 American politicians labeled with their political affiliations. Summary statistics indicate a liberal bias in NYT articles, with an average pro-Democrat score of 0.6 (where 1 represents full pro-Democrat alignment

[4]We focus on the NYT and WSJ, not only because they are major newspapers, but also because they represent contrasting political ideologies, with the NYT seen as liberal and the WSJ as conservative. This enables a balanced analysis of ideological influences on journalistic practices. Our work complements prior work which has studied local newspaper bias using very different identification strategies (Gentzkow and Shapiro, 2010), and when possible we show that key variables like percent of reader visits starting at the homepage is comparable for other national publications. Focusing on the NYT and WSJ has other advantages. Their extensive presence across multiple platforms enhances our comparative media analysis, offering a contrast to the single channel news-feed for other outlets. In particular, we can link the articles appearing on their homepages to their corresponding print editions and podcasts, providing a comprehensive understanding of their editorial strategies, which is not possible with broadcast media like Fox News and CNN.

and 0 indicates complete pro-Republican alignment), while WSJ articles exhibited more neutrality, with an average score of 0.5.

Our primary analysis employs a parametric survival model to explore the duration articles stay on the homepage of newspaper websites. Our findings highlight the importance of demand-side factors, especially the influence of hourly tweets. Articles receiving more tweets per hour consistently show a lower hazard rate, indicating prolonged presence on the homepage. Readers engagement matters.

Even after accounting for the demand-side, supply-side factors significantly influence homepage presence, particularly for the NYT where articles with a liberal slant tend to stay longer. While supply-side influences are generally insignificant for the WSJ, they play a role in specific contexts such as the US section or articles tagged with politically charged keywords like "Politics" where more conservative articles exhibit longer homepage durations after adjusting for the demand-side. We continue to find these effects when we focus on articles which are placed at the very top of the homepage. To account for variation in topics covered, we also limit our sample to articles covering the same news story in both newspapers. We find the same patterns with more liberally-slanted news remaining on the NYT homepage longer, while more conservatively-slanted news staying longer on the WSJ homepage. Notably, this effect is primarily concentrated in the US section, suggesting that supply-side biases are most pronounced in domestic news coverage.

If media bias is mainly driven by supply-side factors like the ideological leanings of journalists and editors, increased competition in a more diverse media market should theoretically reduce bias by promoting accuracy and mitigating one-sided narratives (the opposite is true under models of demand-side bias). This theory suggests that media bias diminishes when demand is more elastic, as outlets compete for a broader, more varied audience, reducing the impact of any inherent bias.

We test this hypothesis by examining two platforms: the print edition of NYT and the flagship podcasts of both NYT and WSJ. The print medium, catering to a stable readership, exhibits more inelastic demand. Our findings indicate stronger biases within specific sections of the print edition, suggesting that inelastic demand enhances the influence of supply-side factors on editorial decisions.

In contrast, podcasts, which the newspapers use to attract new subscribers, face more elastic demand. Our study of podcast editorial decisions reveals that demand-side factors, such as article popularity measured by tweet counts, significantly influence content

selection. However, supply-side factors like pro-Democrat scores showed no significant correlation with podcast feature decisions. This suggests that in environments with increased competition and elastic demand, such as podcasts, media bias is likely less pronounced. In short, these results suggest medias bias is largely supply-driven.

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 Section 5. Section 6 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. Perhaps the central question in this literature is the source of bias. One strand focuses on demand-driven bias: readers have preference for news which affirms their own political views (Mullainathan and Shleifer, 2005). Another set of papers focuses on the supply-side: bias stems from the ideological leanings of the owners, editors or reporters of the media source (Gentzkow and Shapiro, 2006). Our paper seeks to test the relative importance of these two channels. This has real-world importance. Consider the role of competition. Under supply-driven models, competition diminishes bias since consumers will have more choices and they prefer less bias. Conversely, under demand driven bias competition can increase bias through a Hotelling model mechanism: while a monopoly will locate in the political center to maximize readership/profits, while duopolists will symmetrically locate on opposite sides of the political middle to focus on left or right consumer segments.

There is a large and growing literature assessing the extent and sources 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,[5] including demand for media bias (Huang, Meng, and Weng, 2022; Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017, Puglisi and

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

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; Cagé et al., 2022), and the level of market competition 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. Another important advance of our paper is that the methodological approach we use to parse demand- and supply-bias, examining digital homepages, can be applied to virtually any news source.

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 Corniére and Sarvary, 2022; Jeon and Nasr, 2016), the impact of social media on fake news (Allcott and Gentzkow, 2017; Zhuravskaya, Petrova, and Enikolopov, 2020), 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 circu-

lation (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 Corniére 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

Our main results focus on article homepage duration. We show here the homepage shapes what readers view and read. Senior editors understand this and have stated selecting the content of the homepage is one of their highest priorities. Joe Kahn, the NYT's executive editor, said in an interview:

> I think the difference between coming to the Times at any given moment and going to many other places where you can find news on the Internet is we're constantly thinking about the hierarchy of the stories that we're promoting, which is a big part of my day and a big part of my leadership's team day—the play and the prominence of the different story lines that we have ... Multiple times a day, we're debating essentially what the top of the home screen, what the top of the digital feed should look like ... The ranking of those [articles], the packaging of those, is very much a human curation task and an editing task. And we try to put them together really with a lot of intent every day and multiple times a day (Malone, 2024).

The homepage is the most critical digital real estate of a newspaper website, acting as the main access point for articles and multimedia content. A 2011 Pew Research Center study highlights its significance, noting that for the majority of sites (21 out of 25) analyzed, including NYT and WSJ, the homepage is the most visited section (Olmstead, Mitchell, and Rosenstiel, 2011).

Recent data further emphasizes the homepage's crucial role in attracting readers. According to the 2022 annual report from The New York Times Company, their home-

page NYTimes.com received approximately 99 million unique monthly visitors from the United States via desktop and mobile devices. This figure increases to about 145 million globally, demonstrating its extensive digital reach across diverse global audiences (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.[6] 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 confirms the homepage's crucial role for digital news and emphasizes the importance 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)

[6]Direct traffic is similarly important to other media sources. We collected the top ten similar media sites to each of the NYT and WSJ, which SimilarWeb provides based on overlap of their audience, content, keywords and referrals. These lists overlap and include a range of US and non-US print and broadcast sources, such as Fox News, Washington Post, AP, ABC News, the Economist, BBC, the Guardian and Bloomberg. Using all worldwide traffic, the NYT competitors had an average of 44% direct traffic while the WSJ competitors had an average of 43%.

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

Our scraper collects data from the NYT homepage, `https://www.nytimes.com`, and it ran every minute 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,[7] section, tags (keywords), 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.

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 our 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, keywords, 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

[7] The three news tones are "News" (which is mainly hard news), "Opinion" (which is mainly op-ed), and "Feature" (which is mainly soft news).

news tone as in the NYT data. Instead, we categorize an article's news tone by its section which aligns with the NYT sorting.[8]

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

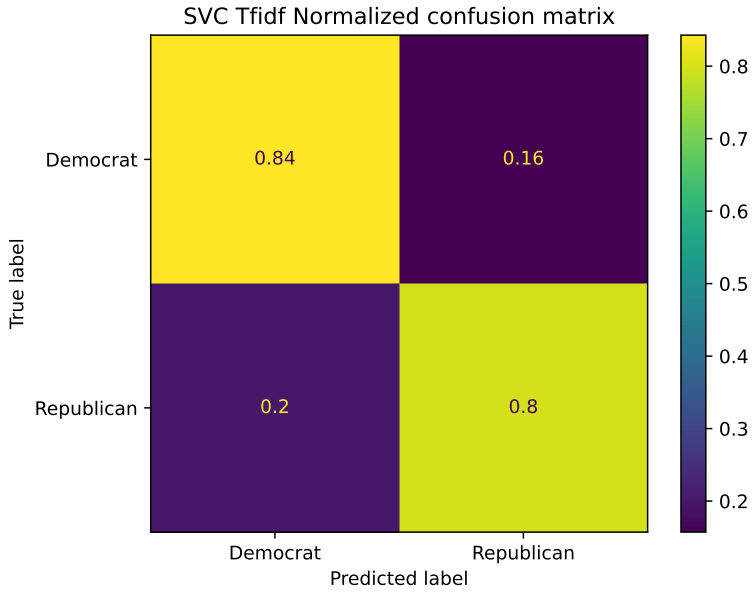
4.1.1 Textual Analysis

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.[9] The first measure is a sentiment score of the text, which ranges from -1 (most negative) to 1 (most positive). 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 media jargon.[10] 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 captures 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.[11] We then use the Twitter API to obtain all their tweets (approximately 2 million tweets in total).

[8]The "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".
[9]Interested readers are referred to the paper for more details on the construction of these measures.
[10]We also use VADER to obtain sentiment scores for tweets which will be discussed in Section 4.2.
[11]We dropped the independent politicians in our sample.

Figure 1: Confusion Matrix of Machine Learning Prediction



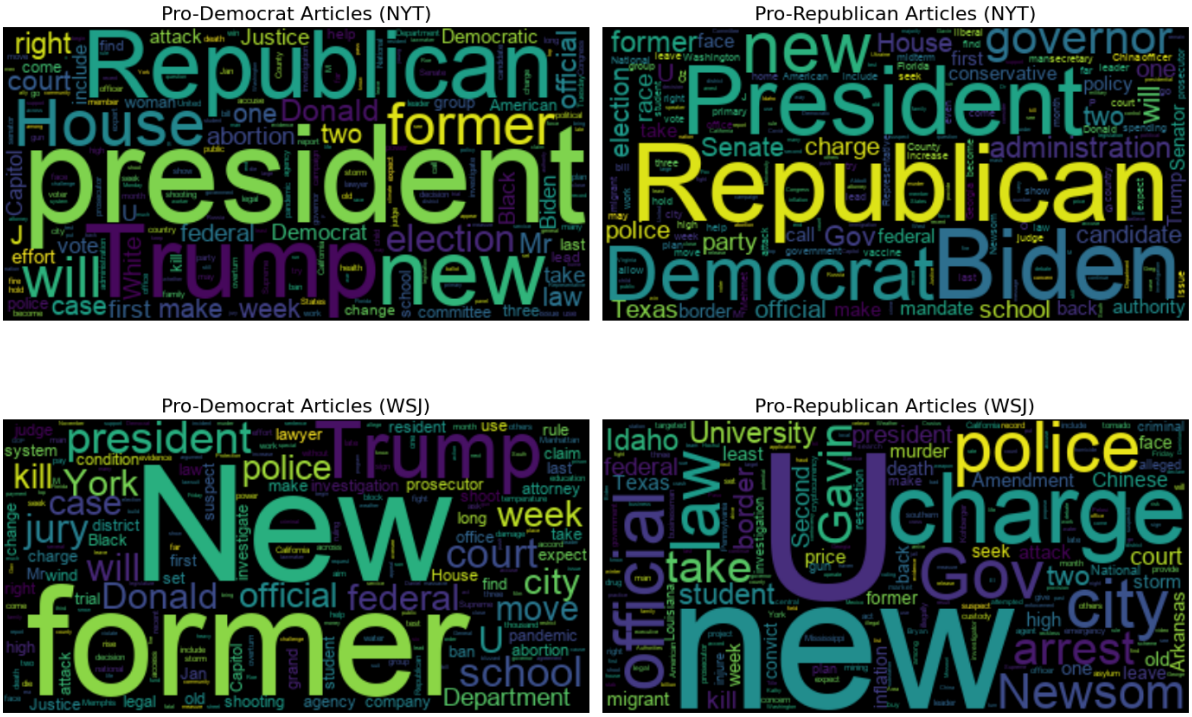
After applying the standard pre-processing steps, we then use the Linear Support Vector Classification (Linear SVC) model in the scikit-learn package, a Python module, as our machine learning model.[12] We fit the model using the training dataset (approximately 1.4 million tweets). Overall, the model achieves 82% accuracy on the leave-out validation dataset (0.6 million tweets). 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 (we presume Democrats are liberal and Republicans conservative). The confidence of the prediction (in terms of probability of it being written by a Democrat ranges between 0 and 1) is our measure of slant.

Figure 2 presents lexical trends within U.S. news articles from the NYT and WSJ, revealing several distinct patterns.[13] The word clouds are divided into two main panels:

[12]We analyze both unigram and bigrams in our machine learning model.
[13]We present more word clouds in other news sections in Appendix A.

Figure 2: Word Clouds of US News Articles' Abstracts



the upper panel displays linguistic trends in NYT articles, with pro-Republican articles (pro-Democrat scores below 0.3) on the left and pro-Democrat articles (pro-Democrat scores above 0.7) on the right. The lower panel shows the most frequent terms within WSJ articles, following the same political score distinctions.

Both the NYT and WSJ pro-Democrat articles exhibit a notable overlap in frequently used terms. Common lexemes such as "former", "President", "Trump", "abortion", and "justice" underscore this shared thematic emphasis.

The pro-Republican articles have more vocabulary variations across the two publications. In the NYT, pro-Republican articles predominantly feature terms such as "Republican", "President", "Biden", and "Democrat". In contrast, WSJ's pro-Republican word choice include "police", "murder", "Newsom", and "migrant".[14] These differences hint at diverging editorial foci and narrative frameworks adopted by the two newspapers.

[14]There are a few terms such as "Texas" and "border" which feature in both NYT and WSJ's pro-Republican articles.

4.1.2 Summary Statistics

Next we overview the newspaper data, starting with statistics on the homepage which is central to our analysis.

Figure 3 displays the homepage duration distributions for NYT and WSJ articles, with the left panel for NYT and the right for WSJ. Notably, 5% of NYT articles and 8% of WSJ articles were removed from the homepage within the first two hours of posting.

Op-ed articles in NYT notably remain on the homepage longer than other categories, with over 15% staying visible for more than 48 hours, compared to less than 3% for other news tones in both NYT and WSJ. This extended visibility suggests that op-eds could have a more substantial impact on readers.

There are spikes at certain durations showing that both NYT and WSJ editors frequently withdraw articles at specific intervals, such as after 6 or 12 hours, particularly for NYT's Feature articles. This pattern indicates a strategic approach to regularly refreshing homepage content, which influences the exposure duration of various 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, 60% of NYT articles and 70% of WSJ articles at some point are featured on their homepages. Since most articles reach the homepage, we do not focus on the entry decision in the main analysis.

Secondly, sentiment analysis reveals a tendency for articles on the homepage to exhibit slightly more negative tones than those that have not been featured, consistent across both NYT and WSJ. This suggests that editors might associate negative sentiment with greater newsworthiness.

Thirdly, the political slant scores show that NYT articles are generally more pro-Democrat or liberal, with average scores around 0.6, while WSJ articles display more moderate scores a bit above 0.5. These mean values corroborate the commonly perceived political inclinations of these two major publications.

Lastly, within the same publication, political slant is roughly the same for articles that appear or do not appear on the NYT homepage. There are more pronounced, but still small, differences in the WSJ, where homepage articles are slightly more liberal than those not featured.

4.2 Twitter API

Ideally, we would assess article popularity using actual view and share data from NYT and WSJ, but this information is unavailable. Instead, we approximate reader demand by counting the number of tweets linking to an article's URL.

Twitter serves as a significant news source, according to a 2021 Pew Research Center survey of 2,548 users (Anderson, Rainie, and Nolan, 2021). While 23% of Americans use Twitter, about 69% of U.S. Twitter users report getting news from the platform. Despite general skepticism towards social media information—only 27% of Americans trust it—two-thirds of Twitter news consumers express some level of trust in the news they get from Twitter. The reliance on Twitter for news has grown; 70% of Twitter news consumers now follow live events on the site, an increase from 59% in 2015.

Between January 2021 and June 2023, Twitter made its entire tweet archive accessible to academic researchers. Using the Twitter API, we extracted all tweets containing URL links to NYT and WSJ articles in our sample. For each tweet, we gathered content, date and time, likes, retweets, and author information such as name, and Twitter ID. We also calculated sentiment scores for each tweet similarly to the article abstracts. Our database includes over 24 million tweets: 22 million for the NYT sample (approximately 275 tweets per article) and 2 million for the WSJ sample (approximately 100 tweets per article).

Figure 4 illustrates the interaction between social media activity (measured in tweet counts) and an article's presence on a newspaper's homepage. It compares two articles covering the same-sex marriage vote mentioned in the introduction: NYT's "Here Are the 12 Republicans Who Voted to Protect Same-Sex Marriage" and WSJ's "Senate Votes to Advance Same-Sex Marriage Legislation."

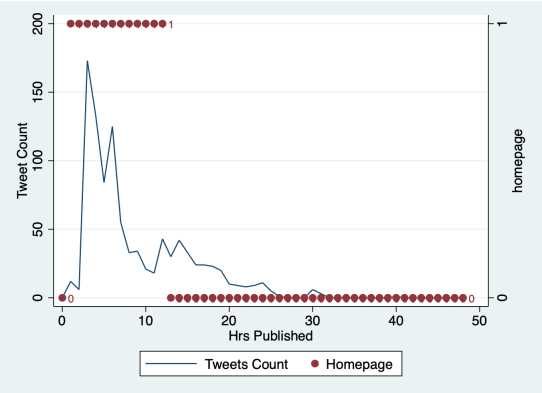
Both articles had a liberal-leaning tone, with pro-Democrat scores of 0.62 for the NYT and 0.71 for the WSJ. Social media engagement showed the NYT article garnered 533 tweet shares in the first six hours, peaking at nearly 180 tweets in the fourth hour, while the WSJ article attracted 169 shares, peaking at about 60 tweets in the fourth hour. Adjusting for typical engagement levels—NYT articles generally receive three times more tweet shares—the engagement for both articles was comparable.

However, the homepage duration differed significantly: the NYT article remained on the homepage for 12 hours, while the WSJ article was featured for only 4 hours.

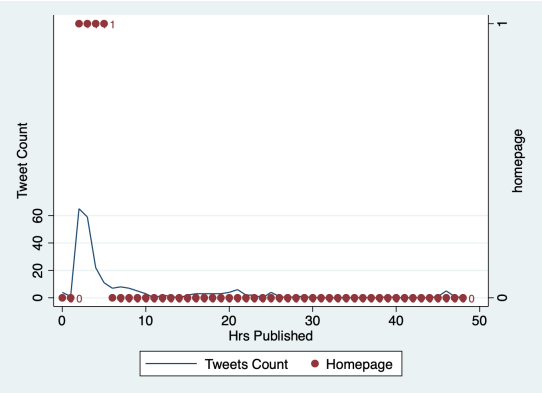
These contrasting trajectories suggest supply-side factors may shape media bias, particularly regarding when a newspaper removes an article from its homepage. While demand-side factors like social media activity play a role, they do not fully explain homepage duration. Thus, it is important to investigate the impact of the media outlet's political inclinations on these editorial decisions.

Figure 4: A Tale of Two Same-Sex Marriage Articles

"Here Are the 12 Republicans Who Voted to Protect Same-Sex Marriage" (NYT)



"Senate Votes to Advance Same-Sex Marriage Legislation" (WSJ)



4.2.1 Validation of Tweets as Measure of Reader Engagement

We use the limited, but direct, NYT analytics to validate that tweets are an effective proxy for reader engagement with an article. Table 3 shows the average hourly tweet count for articles in the top 20 positions in terms of reader views, shares and emails. There are significant tweet differences across the tiers of demand-side popularity of articles.

Table 3: NYT Articles' Rank and Hourly Tweets Count

	Views	Share	Email
Outside Top 20	4.89	4.54	5.01
	(46.66)	(40.90)	(45.53)
Rank 11-20	13.20	16.03	12.33
	(71.65)	(114.8)	(92.24)
Rank Top 10	14.96	24.32	17.01
	(66.20)	(48.25)	(87.11)

Mean coefficients; SD in parentheses

Articles not in the top 20 have hourly tweet counts ranging from 4.5 to 5, while top 20 articles average between 12 and 24 tweets per hour. Within the top 20, top 10 articles enjoy even higher tweet counts compared to those ranked 11-20. This disparity is most evident in share rankings, with top 10 articles averaging 24 tweets per hour versus 16 for the 11-20 group. A similar trend, though less pronounced, is seen in email rankings, with top 10 articles averaging 17 tweets per hour compared to 12 for those ranked 11-20.

Overall, these findings validate the use of tweet counts as a reliable measure of an article's popularity, supporting its application in our analysis.

4.2.2 Summary Statistics and Homepage's Causal Impact on Tweets

Table 4 provides a quantitative summary of Twitter engagement for articles from the NYT and WSJ, aggregating activity across each article's lifetime.

Table 4: Summary of News Articles' Tweets Count

	NYT		WSJ	
	Homepage (No)	Homepage (Yes)	Homepage (No)	Homepage (Yes)
Tweets Count	69.0	409.6	55.8	141.6
	(559.9)	(1510.7)	(247.4)	(1042.9)
Tweets % (Neg)	0.0581	0.0778	0.0730	0.0877
	(0.0890)	(0.0788)	(0.0933)	(0.0891)
Tweets % (Pos)	0.0832	0.0833	0.0985	0.0977
	(0.0993)	(0.0680)	(0.0994)	(0.0819)
N	33191	48501	6357	14250

Mean coefficients; SD in parentheses

First, the data shows that NYT articles are more than twice as popular on Twitter as WSJ articles, with average tweet counts of 277 and 116, respectively. For articles featured on their respective homepages, the gap widens: NYT articles receive about three times as many tweets (410 on average) compared to WSJ articles (142 on average).

Second, homepage articles attract significantly more tweets for both publications. NYT articles on the homepage garner about six times more tweets, while WSJ articles see a threefold increase. This suggests that homepage placement either amplifies an article's visibility and popularity or that more newsworthy articles are featured on the homepage—or both. Appendix B presents estimates of the causal impact of homepage presence on tweet count, using time-of-day homepage updating patterns as an instrumental variable. There are distinctive hours of the day in which there are large purges of homepage content (likely associated with editor shift changes). These hours occur throughout the day and so are unrelated to article newsworthiness though they do shape which articles are on the homepage. After instrumenting we find that tweet counts for articles increased significantly when featured on the homepage (35% increase for NYT articles and 162% for WSJ articles).

Third, 14-18% of tweets for articles from both newspapers express a discernible sentiment—either positive or negative—with a slight preference for positive sentiment. This indicates a relatively balanced sentiment distribution in reader responses.[15]

5 Homepage Survival Analysis: Model and Results

Our main analysis focuses on article homepage duration. The previous sections have shown articles on the homepage get more reader engagement, likely because this is where a majority of readers start their website visit. Senior editors are aware of this and spend significant times curating the homepage. Extended homepage duration can be a sign of bias, and in this section we measure this duration and partition it between demand- and supply-driven channels.[16]

We estimate 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 supply and demand side factors — variables such as sentiment, political slant, and tweet counts— as well as other factors which may influence homepage longevity like timing of other new articles (more newsworthy articles may crowd-out those currently 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:

$$h(\tau; X_{it}, \beta, \lambda, \alpha) = \alpha\lambda(\tau\lambda)^{\alpha-1} \exp(\beta X_{it}), \tag{1}$$

where τ 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 at time t, including the natural log of zero-inflated

[15] Following Hutto and Gilbert (2014) we score each tweet based on the proportion of text which receives a positive, negative, or neutral score. The percent values in the table are the average value for the positive and negative values across all tweets. Note that the overall tweet sentiment is a compound score based on the summed valence scores of each word in the lexicon.

[16] We do not focus on the entry decision, since most articles reach the homepage at some point. We also showed earlier there is little difference in political slant between articles that do and do not get on the homepage.

tweets at $t-1$,[17] sentiment and pro-Democrat scores of the article, and the number of new articles published by the newspaper at time t.[18] 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).

In this study, we are particularly interested in distinguishing the demand and supply side factors contributing to media bias, a division which the model accommodates.

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. Sections 4.2.1-4.2.2 and Appendix B provide evidence that tweets capture reader demand.[19]

Secondly, on the *supply side*, we measure an article's political slant using pro-Democrat scores derived from the article text. While NYT is generally more liberal and WSJ more conservative, individual articles within each newspaper vary widely in their political slant. By incorporating pro-Democrat scores, we can determine if articles with a particular leaning remain on the homepage longer. If supply-side factors, such as editorial bias, are significant, we might expect articles that align more closely with the newspaper's political stance to have extended homepage durations. Key to our approach is the simultaneous control of for the demand-side via tweet counts.

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.

5.1 Main Results

In Table 5, we present our survival analysis results separately for NYT (left panel) and WSJ (right panel). There are estimates for the paper overall as well as by news tone (the left-most columns for each paper). The results are shown as hazard ratios, which measure the effect of each covariate on the likelihood of an article being removed from

[17] Zero-inflated tweet counts is the number of tweet counts plus one.

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

[19] One concern is that tweet counts embed both reader interest and (via the results in Appendix B) the boost from being on the homepage. However, the duration model only includes articles already on the homepage, so the latter effect is a constant term which drops out of the hazard.

the homepage. A hazard ratio less than one indicates that as the covariate increases, the article is less likely to be removed, meaning it stays on the homepage longer.

Table 5: Survival Analysis of Homepage (Hazard Ratios)

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

p-stats are for hazard ratio differences from one.

Exponentiated coefficients; Standard errors in parentheses

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

A key finding from our analysis is the significant impact of demand-side factors. There is a strong negative relationship between tweet counts at $t-1$ (a proxy for article popularity) and the hazard rate (the likelihood of an article being removed from the homepage). For the NYT, a one standard deviation increase in log tweets (1.37) leads to 2.5% decrease in the hazard rate. This effect is even more pronounced for WSJ, where a one standard deviation in log tweets (1.09) leads to a 10.0% decrease in the hazard rate. This indicates that more popular articles, as measured by tweet counts, tend to remain on the homepage longer, underscoring the role of reader demand in determining homepage duration. This result also holds for different news tones and is robust to a variety of changes.[20]

Supply-side factors are particularly influential for the NYT. A one-standard-deviation increase in an article's pro-Democrat score, indicating its political slant, is associated with a 2% decrease in the hazard rate. This relationship holds across different news tones, suggesting that a wide range of NYT's editorial decisions on homepage duration are influenced by the article's political alignment. The results are also robust to alternative interpretations of the supply-side variable.[21]

[20]It is possible some tweets are criticizing an article rather than reflecting demand to read it. To address this potential concern, we conducted a robustness check by replacing the total number of tweets with the count of positive sentiment tweets only. Results reported in Appendix C remain largely similar. Also we find a similar effect if we use as a covariate tweets without the log.

[21]There might be heterogeneity in the duration of article popularity among readers based on its parti-

For the WSJ, the influence of supply-side factors is generally negligible, both overall and for each news tone. The pro-Democrat scores of articles do not significantly affect their duration on the homepage, contrasting with the NYT, where political slant plays a noticeable role. This suggests that the WSJ might prioritize profit-maximizing objectives over ideological considerations (recall demand-side effects are also larger for the WSJ). However, evidence of supply-driven media bias is found in specific sections, such as the US section, and in articles tagged with politically charged keywords like "Politics" (See Appendix E, which also shows the longer duration of liberal articles on the NYT holds by section and keyword) .

While not central to our analysis, the remaining covariates shed light on editorial decision-making. The sentiment of an article, measured by abstract sentiment scores, is strongly linked to its duration on the homepage. For the NYT, a one standard deviation increase in sentiment scores results in a 2% decrease in the hazard rate, meaning more positively toned articles stay longer. This effect is even more pronounced for the WSJ, where a similar increase corresponds to a 4% decrease in the hazard rate.

Additionally, the publication of new articles within a given hour increases the hazard rate, suggesting that new content accelerates the removal of existing articles from the homepage. This is not a mechanical relationship: unlike a print newspaper, there is no fixed space constraint on the digital homepage, allowing NYT and WSJ to vary the number of articles over time.

All of these results extend to the positioning of articles *within* the homepage (Appendix F has details). The top-most articles are the first ones a reader encounters and so are the most prominent. We can extract an article's homepage position using its API ordering or URL suffix. All of the effects above continue to hold when we restrict our focus to article duration on the top section of the homepage. In particular we continue to see evidence of supply-side bias: at the keyword-level there are partisan effects in both papers even when controlling for tweet count.

sanship (e.g. NYT readers have a more sustained interest in pro-Democrat articles). In this case more liberal articles on the NYT would see a slower decay in tweet count, and the same with conservative ones on the WSJ. Editors may therefore keep more partisan articles on the homepage longer to satisfy anticipated reader demand rather than due to supply-side bias. However, Appendix D demonstrates that articles, regardless of their liberal or conservative tone, exhibit a consistent demand decay trajectory.

5.2 Article Survival for Identical News on NYT and WSJ

One concern with the survival analysis in the last sub-section is that the news topics can differ between the papers. Here we consider homepage duration for matched articles covering the same story. By analyzing the survival rates of articles covering identical news stories in both newspapers, we aim to uncover how editorial decisions influence the homepage endurance of the same event. This approach examines whether liberal-leaning articles remain longer on the NYT homepage and likewise with conservative-leaning articles on the WSJ homepage.

To ensure a precise comparison, we identified pairs of articles from each newspaper that reported the same news on the same day by calculating the cosine similarity between their headlines and abstracts (articles can be part of multiple news topic clusters and so matched to more than one article from the other paper).[22] Articles were considered to cover the same news if their cosine similarity score exceeded 0.2.[23]

During our sample period from October 2022 to May 2023, we identified 2,031 pairs of news articles that appeared on both newspapers' homepages. This cohort includes 1,817 unique NYT articles (10.4% of 17,452 sampled) and 1,706 unique WSJ articles (12.0% of 14,250 sampled). In our matched sample over 90% of the NYT articles and 80% of the WSJ articles are categorized as News articles, instead of Feature or Op-Ed articles. Figure 5 shows that within this News category, the predominant sections are Business, US, and World, collectively encompassing more than 70% of the identified articles.

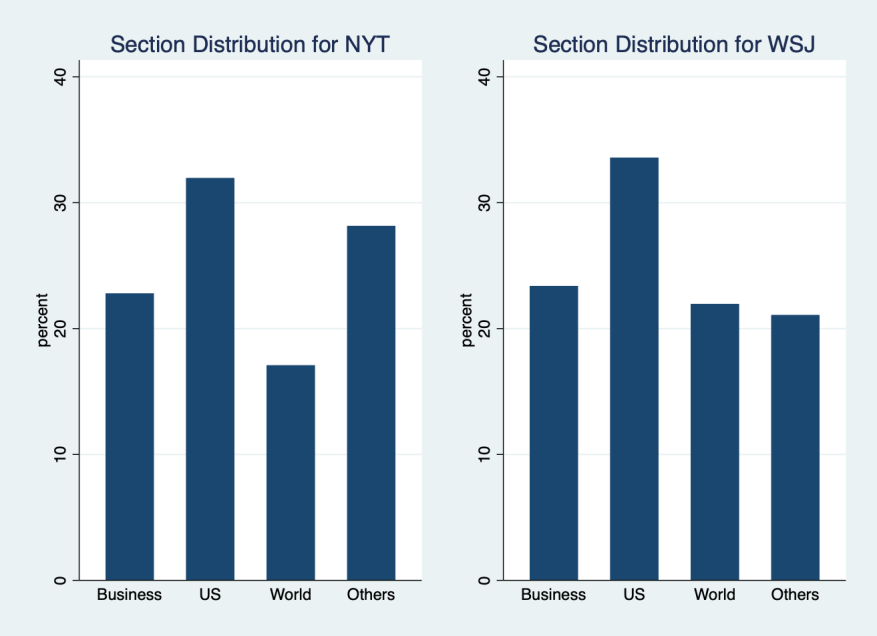
We used a simple Ordinary Least Squares (OLS) regression to analyze the difference in survival length (hours) on the homepages of NYT and WSJ. The model included variables such as the pro-Democrat scores of the articles, the difference in tweet counts within the first six hours of publication, and other time-fixed effects. Since pro-Democrat scores often differ between the two newspapers even for the same news, we consider different specifications which either average pro-Democrat scores from the articles in the two papers or use specific scores from either the NYT or WSJ. We also have specifications which restrict our sample to article pairs with close pro-Democrat scores.

[22] Text preprocessing included lemmatization to group similar words and removal of stop words for enhanced matching accuracy.

[23] Similar results were obtained with thresholds of 0.25 and 0.3, confirming the robustness of our matching criterion.

To more formally validate our approach, we had a research assistant manually match similar articles from the two newspapers on three dates. The manually matched sample closely resembled our sample obtained through textual analysis. In fact, our sample had fewer false positives.

Figure 5: Distribution of Sections Among Matched News Articles



The results in Table 6 reveal two primary findings. First, tweet count differences significantly influence survival duration; articles with more NYT- than WSJ-related tweets remain longer on the NYT homepage. Second, even after adjusting for tweet count differentials, more liberal articles consistently show extended survival on the NYT homepage compared to WSJ. For example, a one standard deviation increase in average pro-Democrat scores (column one) correlates with an additional hour on the NYT homepage. This pattern holds across various specifications, including average pro-Democrat scores (column one), individual scores from each newspaper (columns 2-3), and results are stronger (in terms of economic magnitude) among articles with closely matched pro-Democrat scores (columns 4-5).

Table 6: OLS Regression of Survival Difference (NYT - WSJ)

	(1) Avg Pro-Dem Score	(2) NYT Pro-Dem Score	(3) WSJ Pro-Dem Score	(4) Pro-Dem Dif. < 0.1	(5) Pro-Dem Dif. < 0.05
Pro-Dem. Scores	4.791***	2.424*	3.958***	4.918*	6.830*
(Abstract)	(1.653)	(1.334)	(1.360)	(2.574)	(3.940)
Tweets Dif.	0.000886***	0.000904***	0.000897***	0.000953**	0.00128*
(NYT - WSJ)	(0.000250)	(0.000250)	(0.000250)	(0.000424)	(0.000687)
Observations	2031	2031	2031	626	306
Day FE	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses

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

Building on the analysis in Table 6, Table 7 focuses on each section. There is a pronounced supply-side bias in the U.S. news section. A one-standard deviation increase in average pro-Democrat scores is associated with an additional 1.13 hours on the NYT homepage, highlighting a preference for liberal slants in this section. In contrast, articles in the Business and World sections do not show this trend; more liberal articles there do not have prolonged survival on the NYT homepage compared to their WSJ counterparts.

Table 7: OLS Regression of Survival Difference by Sections (NYT - WSJ)

	(1) Business	(2) US	(3) World	(4) All
Pro-Dem. Scores	-1.658	6.293**	-1.585	4.791***
(Abstract)	(4.005)	(2.747)	(3.751)	(1.653)
Tweets Dif.	0.00348***	0.000650**	0.00107*	0.000886***
(NYT - WSJ)	(0.00127)	(0.000313)	(0.000646)	(0.000250)
Observations	339	546	321	2031
Day FE	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes

Standard errors in parentheses

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

6 Editorial Decisions on Print Edition and Podcasts

In this section we test a direct implication of the supply-side model of media bias. Gentzkow and Shapiro (2006) argue that competition can reduce media bias and improve reporting accuracy, as diverse media outlets provide readers with various follow-up stories, facilitating the verification of initial reports. This idea is further explored by Gentzkow, Shapiro, and Stone (2015) and Puglisi and Snyder (2015), who suggest that competition generally reduces distortions, aligning media outcomes more closely with a bias-free state.[24]

A key observation from these insights is that media bias might be influenced by the elasticity of demand. Media facing inelastic demand might exhibit greater bias, while those facing elastic demand might have less bias. This distinction underpins our examination of editorial decisions in 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, likely faces inelastic demand. In contrast, podcasts, aiming to capture new

[24]In contrast, Mullainathan and Shleifer (2005) model bias as demand-driven and find competition increases bias.

audiences, operate in an environment with more elastic demand. Comparing editorial choices across these two platforms, which have similar content but face different degrees of demand elasticity, is a natural test of the supply-side model of media bias.

6.1 Print Edition

The readership of print editions for major newspapers like The NYT and WSJ exhibits notably inelastic demand characteristics. This inelasticity reflects high loyalty among readers who consistently prefer their chosen publication over alternatives, despite steep price changes or the availability of similar information online.[25]

Empirical evidence supports this. According to Pew Research Center (2016), print newspaper readers tend to be older and more affluent, demographics that correlate with higher brand loyalty and less price sensitivity compared to younger, more digital-centric audiences. These readers value in-depth analysis and comprehensive reporting, making them resistant to switching to other media even as print subscription costs rise.

American Press Institute (2018) also highlights that print subscribers have strong emotional and historical ties to the newspaper. The report notes that print subscribers are significantly more likely to cite a longstanding relationship or fond memories with the paper as reasons for their loyalty compared to other subscribers (19% versus 2%). This connection underscores the inelastic nature of their demand and their resilience in the face of shifts toward digital media formats.

In examining editorial decisions regarding print edition articles, we focus on the NYT due to data availability constraints.[26] We used the presence of a "print headline" from the NYT API as an indicator of an article's appearance in the print edition. We manually verified the "print headline" from the API against the one listed at the bottom of articles on the NYT website. This variable accurately represents an article's inclusion in the print edition up to February 2023. Historically, 60% to 70% of articles per month appeared in the print edition, but starting February 2023, this figure dropped to below 40%. To address this anomaly, we limited our analysis to articles up to January 2023, resulting in

[25]The cover price of the print weekday NYT increased from \$0.75 to \$4.00 between January 2000 and January 2024. If its price had increased at the same rate as inflation, its price in 2024 would be \$1.37.

[26]WSJ articles do not list whether they are in the print edition, unlike the NYT. There is an archive of WSJ print editions available at `https://www.wsj.com/print-edition/today`, but it is only available for the prior three months. We contacted Dow Jones, the WSJ publisher, but they were unable to generate a print archive for our sample period.

a sample size of 63,740 articles.

Table 8 presents estimates of whether an article is included in each of the major sections of the print edition (Appendix G has analogous estimates by news tone as well as overall). Using a linear regression model with the binary outcome of print inclusion as the dependent variable, we find that articles featured on the homepage or those with a high number of tweets are more likely to be selected for the print edition. Even with these controls there is evidence of supply-side bias: articles with more pro-Democrat scores are more likely to be added to the NYT print edition. This is particularly evident in prominent sections like "US", "Business", and the feature sections "Arts" and "Style". For instance, in the "Arts" section, a one-standard deviation increase in an article's pro-Democrat score correlates with a 3.4% higher probability of appearing in the print edition.[27]

The strong effects in this environment of limited competition are consistent with the predictions of supply-side theories of bias.

Table 8: NYT Print Edition Linear Regression by Major Sections

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

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

6.2 Podcasts

The rapidly increasing popularity of podcasts as a news medium is notable. Pew Research Center (2023) finds that about half of Americans listened to a podcast in 2022, with many encountering news content. Meredith Kopit Levien, Chief Executive of The New York

[27] Two sections have non-significant parameters. "Life" groups several smaller Feature print sections (including "dining", "sports","travel", "magazine", and "well"), and "World" is likely to face relatively more competition from other news sources.

Times, highlighted this shift during the Times' third-quarter results in 2020. Levien noted the importance of The Daily, the newspaper's flagship podcast, which has an audience of 4 million daily, nearly twice the paper's peak circulation (Turvill, 2020).

Levien also highlighted the demographic shift in this audience, with most listeners under 50 and many under 40. The podcasts cultivate greater loyalty among this younger demographic who develop "an affinity for The New York Times." This change is important for the paper, since younger readers disproportionately use news aggregators like Google News or Facebook News Feed. Reuters Institute (2023) notes that younger users are less likely to visit a news site directly and more likely to use social media or other intermediaries, potentially reducing loyalty to any single news source.

Podcasts not only attract a large audience but also play a crucial role in solidifying brand loyalty and boosting newspaper subscriptions. Although Levien acknowledges challenges in directly attributing subscription growth to the podcast, she remains optimistic, stating, "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."

We focus on understanding the editorial decisions driving article selection for flagship podcasts. Specifically, we analyze articles featured in The Daily by the NYT and The Journal by the WSJ during our sample period. To gather this data, we visited the URLs of the respective NYT and WSJ podcasts daily and extracted the URLs of the "feature articles" presented in each episode. These articles were then matched to our primary dataset. Only articles featured in a podcast within 30 days of their original publication were included in our analysis.[28]

Sample statistics in Table 9 reveal notable trends in article selection for inclusion in the newspapers' podcasts. Only about one percent of articles are chosen for the podcasts. This is due to the operational logistics of producing daily episodes that typically feature up to three articles, while the newspapers publish over 100 articles daily. Podcasts are also primarily based on News articles, and none are from Opinion articles.

Articles featured in the podcasts are far more popular with readers and receive extra promotion. For both newspapers, tweet counts within the first 12 hours post-publication for these articles are about five times higher than for those not featured, highlighting greater reader engagement with podcast-selected content. Additionally, articles featured in podcasts are more likely to have been displayed on the newspapers' homepages, with

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

a nearly 30 percentage point higher likelihood compared to non-featured articles.

In terms of sentiment, articles chosen for podcasts generally have a more negative tone. Critical or emotionally intense content is more likely to attract podcast discussion, potentially due to its capacity to provoke deeper analysis or listener interest.

Most importantly for our focus, there is no significant difference in pro-Democrat scores between articles featured in podcasts and those not featured. This political neutrality supports the hypothesis that the podcast selection process, encountering more elastic demand, does not strongly prioritize articles based on political slant.

Table 9: Summary of News Articles by Podcast Status

	NYT		WSJ	
	Podcast (No)	Podcast (Yes)	Podcast (No)	Podcast (Yes)
Ever on Homepage	0.603 (0.489)	0.882 (0.323)	0.689 (0.463)	0.941 (0.235)
Tweets Count	151.9 (731.9)	725.4 (2407.8)	52.59 (238.9)	270.3 (505.2)
Sentiment Scores (Abstract)	0.0199 (0.432)	-0.128 (0.471)	0.0227 (0.412)	-0.0579 (0.405)
Pro-Dem. Scores (Abstract)	0.600 (0.211)	0.601 (0.215)	0.533 (0.211)	0.529 (0.217)
tone (News)	0.573 (0.495)	0.944 (0.230)	0.501 (0.500)	0.802 (0.400)
tone (Opinion)	0.0849 (0.279)	0 (0)		
tone (Feature)	0.341 (0.474)	0.0525 (0.223)	0.238 (0.426)	0.176 (0.381)
N	81080	612	20383	224

Mean coefficients; SD in parentheses

Building on these data, Table 10 presents estimates of a linear regression model similar to the print edition specification. Separate results are listed for each news tone as well as overall (there is no Opinion column, since no podcasts are based on an Op-Ed). The conclusions discussed below also hold for major news sections (see Appendix G for details).

Table 10: Podcast Linear Regression by News Tones

	NYT			WSJ		
	News	Feature	All	News	Feature	All
Ever on	0.000801	0.000355	0.000741	0.000138	-0.00305	-0.00275
Homepage	(0.00137)	(0.000561)	(0.000818)	(0.00413)	(0.00334)	(0.00238)
ln Tweets	0.00670***	0.00165***	0.00510***	0.0143***	0.0116***	0.00926***
	(0.000403)	(0.000234)	(0.000260)	(0.00108)	(0.00109)	(0.000628)
Sentiment Scores	-0.00353***	-0.000792	-0.00275***	-0.00454	-0.000527	-0.00307
(Abstract)	(0.00124)	(0.000614)	(0.000776)	(0.00303)	(0.00333)	(0.00201)
Pro-Dem. Scores	-0.000248	-0.00126	-0.000293	0.00383	-0.0109	-0.000277
(Abstract)	(0.00247)	(0.00127)	(0.00155)	(0.00603)	(0.00679)	(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

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

The left panel presents results for the NYT, and the right panel for the WSJ, and there are consistent patterns across all the columns.[29] Demand-side popularity, as measured by tweet count, positively correlates with an article's likelihood of being featured in a podcast. In contrast supply-side factors, such as pro-Democrat scores, do not significantly influence podcast feature decisions. The difference of the latter result from those for the print edition and for homepage inclusion is consistent with the prediction of supply-side models: more elastic demand leads to less media bias.

7 Conclusion

This study introduces a novel methodology for identifying supply-side media bias without requiring detailed knowledge of the ideological preferences of newspaper owners or editors. Using a comprehensive dataset of over 100,000 articles from The NYT and WSJ, this research leverages content from two of the US's largest newspapers, positioned on opposite ends of the political spectrum. Complementary Twitter data control for demand-side influences, enhancing the robustness of our analysis.

By adjusting for demand-side preferences, we attribute the longer homepage presence of ideologically slanted articles to supply-side biases. Estimates show that liberal articles tend to have extended exposure on the NYT homepage, whereas conservative articles

[29]This includes covariates which are not central to our analysis. There is no significant connection between articles being displayed on the homepage and their selection for podcasts. However, articles with a more negative tone are more likely to be featured in The Daily, the NYT's flagship podcast.

maintain a prolonged presence on the WSJ homepage in prominent sections and for politically charged keywords. Further analyses of article placement in the print edition and podcasts, and comparisons of homepage duration for identical news stories across both newspapers, support the presence of supply-side biases.

This study has limitations. We focus on two national newspapers. Still these are the most prominent US newspapers and prior work has focused on local media. Another concern is our presumption that the differing political slants of the newspapers we study are exogenous. The latter may not be an import caveat for two reasons. First, all media sources face the same flow of news, and it is their choice of what and how to report it which is our focus. Second, news events are relatively random which provides an exogenous source of variation in the potential political tone of articles.

Despite these limits, we believe this work provides avenues for future research and has implications for policy-makers. We provide evidence supporting a key role for supply-driven models of media bias which relies on few assumptions and readily available data. Our approach can be applied to other media sources or time periods (digital scans of historical newspapers are becoming more readily available). An important topic to investigate would be the impart of the declining number of newspapers, especially concentrated among local papers, over the last several decades (Abernathy, 2020). While much of the policy discussion of this decline has focused on newspapers' role as government watchdogs and covering local news, it may also have implications for the bias of the newspapers that still remain. While even moderate sized cities used to have multiple papers, now most have one. Policy-makers may have another reason to be concerned about this secular decline in papers if it leads to increased media bias.

References

Abernathy, Penelope Muse. 2020. "News Deserts and Ghost Newspapers: Will Local News Survive?" URL https://www.usnewsdeserts.com/.

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.

American Press Institute. 2018. "Print Fans subscribe because they love relaxing with a paper." URL https://americanpressinstitute.org/print-fans/.

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.

Astor, Maggie. 2022. "Here Are the 12 Republicans Who Voted to Protect Same-Sex Marriage." *The New York Times* URL https://www.nytimes.com/2022/11/16/us/elections/republicans-same-sex-marriage-vote.html. Accessed: date-of-access.

Athey, Susan and Michael Luca. 2019. "Economists (and Economics) in Tech Companiess." *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 Corniére, Alexandre and Miklos Sarvary. 2022. "Social Media and News: Content Bundling and News Quality." *Management Science* 69 (1):162–178.

de Corniére, 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.

Farronato, Chiara, Andrey Fradkin, and Alexander MacKay. 2023. "Self-Preferencing at Amazon: Evidence from Search Rankings." *AEA Papers and Proceedings* 113:239–43.

Ferek, Katy Stech. 2022. "Senate Votes to Advance Same-Sex Marriage Legislation." *The Wall Street Journal* URL https://www.wsj.com/articles/senate-vote-on-same-sex-marriage-to-test-level-of-gop-support-11668623694. Accessed: date-of-access.

Folkenflik, David. 2020. "Rancor Erupts In LA Times Newsroom Over Race, Equity And Protest Coverage." *NPR* URL https://www.npr.org/2020/06/15/874530954/rancor-erupts-in-la-times-newsroom-over-race-equity-and-protest-coverage?ft=nprml&f=874530954.

———. 2024. "NPR defends its journalism after senior editor says it has lost the public's trust." *NPR* URL https://www.npr.org/2024/04/09/1243755769/npr-journalist-uri-berliner-trust-diversity.

Gentzkow, Matthew, Edward L. Glaeser, and Claudia Goldin. 2006. "The Rise of the Fourth Estate. How Newspapers Became Informative and Why It Mattered." In *Corruption and Reform: Lessons from America's Economic History*, edited by Edward L. Glaeser and Claudia Goldin, Corruption and Reform: Lessons from America's Economic History. University of Chicago Press, 187–230.

Gentzkow, Matthew and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114 (2):280–316.

———. 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.

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.

Malone, Clare. 2024. "The Culture Wars Inside the New York Times." *New Yorker* URL `https://www.newyorker.com/news/the-new-yorker-interview/the-culture-wars-inside-the-new-york-times-joe-kahn`.

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.

Pew Research Center. 2016. "Around Half of Newspaper Readers Rely Only on Print Edition." URL https://www.pewresearch.org/short-reads/2016/01/06/around-half-of-newspaper-readers-rely-only-on-print-edition/.

———. 2018. "In Western Europe, Public Attitudes Toward News Media More Divided by Populist Views Than Left-Right Ideology." URL https://www.pewresearch.org/wp-content/uploads/sites/20/2018/05/PJ_2018.05.14_Western-Europe_FINAL.pdf.

———. 2020. "Two-thirds of U.S. adults say they've seen their own news sources report facts meant to favor one side." 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-mea

———. 2023. "Podcasts as a Source of News and Information." URL https://www.pewresearch.org/wp-content/uploads/sites/20/2023/04/PJ_2023.04.18_Podcasts_FINAL.pdf.

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.

Reuters Institute. 2023. "Digital News Report 2023." URL https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf.

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

Statista. 2024. "Largest daily newspapers in the United States in the six months to September 2023, by average print circulation." URL https://www.statista.com/statistics/272790/circulation-of-the-biggest-daily-newspapers-in-the-us/.

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-subscr

Zhuravskaya, Ekaterina, Maria Petrova, and Ruben Enikolopov. 2020. "Political Effects of the Internet and Social Media." *Annual Review of Economics* 12:415–438.

Appendix A Word Clouds of Other News Sections

There is a consistent set of words in articles categorized as either liberal or conservative across the newspapers. This gives some confidence in the machine learning algorithm used to generate the scores.

An examination of word clouds for news articles within the 'World' section in Figure A1 reveals distinct patterns related to geopolitical events. Notably, the Ukraine conflict is frequently mentioned in both newspapers. While this theme appears in both liberal and conservative articles, it is more prominent in liberal articles in both newspapers, suggesting a broader focus or specific narrative angle in liberal coverage of the Ukraine situation.

Figure A1: Word Clouds of World News Articles' Abstracts



In relation to China, keywords such as "China", "Chinese", "Beijing", "Xi", and "Communist" appear more frequently in conservative articles across both newspapers, reflecting the hawkish stance against China among Republican politicians.

Analyzing word clouds for 'Business' section articles in Figure A2 reveals similar the-

Figure A2: Word Clouds of Business News Articles' Abstracts



matic trends. The term "China" is a recurrent motif in conservative articles, consistent with observations from the 'World' section. This suggests a sustained conservative focus on China-related issues across both global and business contexts.

Additionally, the term "inflation" is prominent in conservative articles in both newspapers. This reflects Republican concerns about the negative effects of rising interest rates by the Federal Reserve. The focus on inflation in these conservative articles highlights partisan concerns about economic impacts, underscoring the interplay between fiscal policy perspectives and media representation.

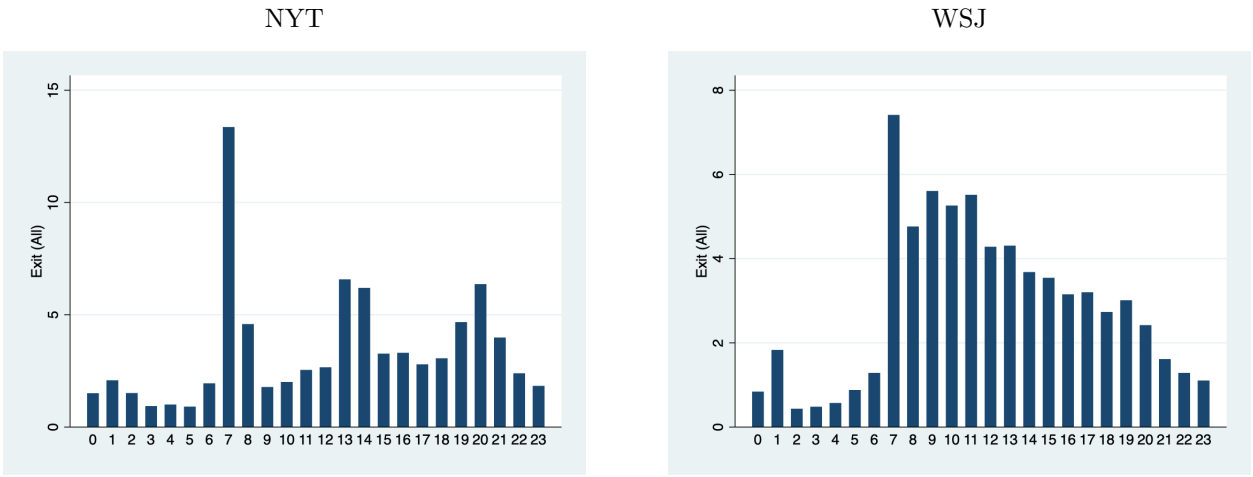
Appendix B Impact of Homepage Presence on Article Popularity (Tweet Count)

One way that the digital homepage matters is through increasing readership of articles.[A1] Establishing such a causal relationship, however, is challenging. A straightforward regression with tweet count (our measure of popularity) as the dependent variable and homepage presence as an independent variable is problematic due to endogeneity. Articles are not randomly placed on the homepage; they are selected based on various unobserved factors, potentially confounding the relationship between homepage presence and popularity.

We address this endogeneity issue using an instrumental variable (IV) approach. Our chosen IV leverages a pattern observed in the timing of article withdrawals from the homepage.

Figure A3 illustrates the average number of articles removed from the homepage by time of day. This patterns, shown for the NYT on the left and the WSJ on the right, highlight certain hours where pronounced homepage updates occur. The spikes are driven by routine editorial practices rather than article demand: they are at times when there is a changeover in editors responsible for the homepage.

Figure A3: Articles Withdrawal by Hours of Day (Eastern Time)



For the NYT, notable spikes in article withdrawal occur at 7am, 8am, 1pm, 2pm, 7pm,

[A1] The homepage, often seen as the digital equivalent of a newspaper's front page, is crucial in shaping readers' perceptions and guiding their attention. Editorial decisions about which stories remain and which are withdrawn reflect inherent biases and significantly influence news dissemination and reception.

and 8pm. For the WSJ, spikes are primarily at 7am and 1am. We use these specific hours as instrumental variables in our analysis to address endogeneity and accurately estimate the impact of homepage presence on article popularity, measured by tweet counts.

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:

$$\ln Y_{it} = \beta_0^H Homepage_{it} + \beta_1^H X_{it} + \phi_i + \varepsilon_{it}, \tag{A1}$$

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 popularity 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 $WithdrawalHours_{it}$ 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 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 in Table A1 analyze the impact of homepage placement on tweet counts

for articles from both newspapers. To maintain consistency with Section 5, our sample includes only articles that were on the homepage in the preceding hour $(t-1)$. This criterion aligns our sample with the main text focus on article homepage longevity.

Table A1: 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***	0.764***	0.263***	0.349***	1.496***	2.070***	1.622***
	(0.0386)	(0.171)	(0.0215)	(0.0241)	(0.144)	(0.319)	(0.131)
Ln Tweets (Lag)	0.407***	0.413***	0.270***	0.400***	0.292***	0.213***	0.278***
	(0.00172)	(0.00230)	(0.00265)	(0.00119)	(0.00403)	(0.00928)	(0.00374)
Hours Published	-0.0268***	-0.0178***	-0.0162***	-0.0207***	-0.0112***	-0.00392*	-0.00916***
	(0.000487)	(0.000497)	(0.000332)	(0.000181)	(0.00118)	(0.00203)	(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***	-0.0190***	-0.170***	-0.0770***	-0.0497***	-0.0432***	-0.0482***
	(0.00109)	(0.000807)	(0.00179)	(0.000706)	(0.00251)	(0.00468)	(0.00221)
Ln Tweets (Lag)	0.0235***	0.00680***	0.0230***	0.0222***	0.0189***	0.0212***	0.0199***
	(0.000563)	(0.000474)	(0.00127)	(0.000394)	(0.000987)	(0.00204)	(0.000890)
Hours Published	-0.0116***	-0.00283***	-0.00983***	-0.00629***	-0.00787***	-0.00616***	-0.00734***
	(0.0000743)	(0.0000331)	(0.000123)	(0.0000382)	(0.0000994)	(0.000163)	(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

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

The table is divided into two panels: the left for NYT results and the right for WSJ. Each panel is further split into two parts: the upper half presents the second-stage results, while the lower half outlines the first-stage findings.

In the first-stage results, both NYT and WSJ show a higher likelihood of article withdrawal during specific hours, consistent with our expectations based on routine editorial practices.

The second-stage results highlight the significant impact of homepage presence on an article's tweet counts. The magnitude varies notably between the two newspapers. For the NYT, homepage presence increases tweet counts by 35% on average, with op-ed articles seeing a 76% rise and feature articles a 26% increase. WSJ articles experience a much larger boost from homepage placement, with an average increase of 162%. News articles

in WSJ see a 150% rise in tweet counts, while feature articles spike by over 200%.[A2]

These estimates show that homepage presence has a robust and economically meaningful impact on an article's popularity as measured by Twitter activity.

[A2] A possible explanation for the smaller impact on NYT articles is the newspaper's effective use of other promotional channels, such as Facebook and Twitter, which may mitigate the relative benefits of homepage placement.

Appendix C Survival Analysis with Positive Tweets

The primary results from our survival analysis, as presented in Table 5, highlight the significant impact of supply-side factors even after accounting for tweet counts, a proxy for demand-side engagement. However, not all tweets may reflect genuine demand for the articles. For example, tweets criticizing the reporting might not correlate with increased views or positive engagement.

To address this concern, we conducted a robustness check by considering only positive tweets instead of the total tweet count (we use our machine learning-based sentiment score to classify articles as positive or not, see footnote 15). The results of this reanalysis, detailed in Table A2, remain consistent with our initial findings. The hazard rate appears even more sensitive to the number of positive tweets than the overall tweet count. Crucially, the conclusions regarding supply-side factors remain unchanged: liberal articles continue to stay longer on the NYT homepage after adjusting for demand-side factors, while the WSJ shows no significant association between supply-side biases and homepage duration.

Table A2: Survival Analysis of Homepage (Hazard Ratios)

	NYT				WSJ		
	News	Opinion	Feature	All	News	Feature	All
Ln Tweets	0.903***	0.540***	1.250***	0.925***	0.734***	0.604***	0.702***
	(0.00809)	(0.0227)	(0.0267)	(0.00752)	(0.0226)	(0.0395)	(0.0193)
Sentiment Scores	0.958***	0.967	0.984	0.961***	0.886***	0.999	0.921***
(Abstract)	(0.0112)	(0.0320)	(0.0202)	(0.00931)	(0.0204)	(0.0394)	(0.0178)
Pro-Dem. Scores	0.916***	0.878**	0.876***	0.900***	0.987	0.950	0.955
(Abstract)	(0.0227)	(0.0531)	(0.0390)	(0.0183)	(0.0472)	(0.0800)	(0.0388)
# of New	1.042***	1.257***	1.015***	1.023***	1.073***	1.044***	1.041***
Articles	(0.00198)	(0.00917)	(0.00339)	(0.00103)	(0.00436)	(0.00779)	(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

p-stats are for hazard ratio differences from one.

Exponentiated coefficients; Standard errors in parentheses

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

Appendix D Does Demand Decay Vary by Political Slant?

In our analysis of homepage duration in Table 5, there is a potential challenge to our interpretation. One might argue that the effects attributed to supply-side factors, measured by an article's pro-Democrat score, could actually be driven by latent demand dynamics. For instance, articles with a pronounced liberal slant (in the case of NYT) or conservative slant (in the case of WSJ) might inherently attract greater sustained interest, leading to a slower decay in demand, even when prior tweet counts are equivalent. In this case our partisan variable need not reflect supply-side preferences but rather editor forecasts of future demand.

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

$$\ln Y_{it} - \ln Y_{it-1} = \beta Z_{it} + \epsilon_{it}, \tag{A2}$$

where Y_{it} is the zero-inflated tweets count for article i at time t, and $\ln Y_{it} - \ln Y_{it-1}$ is the log difference of tweets count, e.g. the decay in tweets. 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, presence on the homepage, 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 A3 indicate that, regardless of the newspaper, tweets for most sections' articles decline by 5.2% to 7.6% per hour in the initial stages after publication, with NYT op-ed articles as the only exception. Furthermore, as articles age, the rate of decay slows, as shown by the estimates on hours since publication.[A3]

[A3] Additionally, sentiment scores are not significantly associated with tweet decay. The only exception is NYT news articles, where those with a more positive tone experience a slower decay in tweets. Homepage presence and number of new articles published that period have mixed effects.

Table A3: Log Difference (Tweets) Regressions

	NYT				WSJ		
	News	Opinion	Feature	All	News	Feature	All
Sentiment Scores	0.00224**	-0.000605	-0.000386	0.00125	-0.000669	0.00204	0.0000995
(Abstract)	(0.00102)	(0.00291)	(0.00122)	(0.000763)	(0.00204)	(0.00268)	(0.00163)
Pro-Dem. Scores	-0.000801	0.00195	0.000572	-0.000214	0.00227	0.00530	0.00319
(Abstract)	(0.00205)	(0.00540)	(0.00252)	(0.00154)	(0.00406)	(0.00541)	(0.00327)
Homepage	0.0197***	-0.0200***	0.0227***	0.0205***	-0.0137***	-0.0139***	-0.0141***
	(0.00134)	(0.00300)	(0.00158)	(0.000952)	(0.00231)	(0.00299)	(0.00183)
Hours Published	0.000850***	-0.00152***	0.00181***	0.00105***	0.00122***	0.000911***	0.00112***
	(0.0000364)	(0.000101)	(0.0000364)	(0.0000253)	(0.0000757)	(0.0000882)	(0.0000580)
# of New	0.00107***	0.0133***	0.00195***	0.000933***	-0.000130	-0.00116**	-0.000825**
Articles	(0.000163)	(0.00105)	(0.000172)	(0.000118)	(0.000415)	(0.000569)	(0.000369)
Constant	-0.0670***	0.00134	-0.0755***	-0.0689***	-0.0742***	-0.0516***	-0.0711***
	(0.00368)	(0.0102)	(0.00410)	(0.00296)	(0.00654)	(0.00808)	(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

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

The most important result is the test of the demand expectations hypothesis. There is little support for this in the estimates: pro-Democrat scores show no statistically significant association with the rate of tweet decay. In other words, articles with a liberal slant in NYT or a conservative slant in WSJ do not exhibit a distinct deceleration in decay compared to other articles. Article partisanship does not impact tweet trajectories. This is consistent with our interpretation in the text that the findings in Table 5 are primarily due to supply-side factors.

Appendix E Survival Analysis by Sections and News Tags

To provide deeper insights into the roles of demand and supply-side factors in determining the duration of articles on newspaper homepages, we divide our samples by major news sections and tags (keywords).

Table A4 presents the results by major news sections, revealing several interesting patterns.

First, higher tweet counts, which represent demand-side factors, typically reduce exit from the homepage for both newspapers. This effect is stronger in the news sections compared to the feature sections.

Second, political slant, which captures supply-side factors, plays a significant role in the NYT as in Table 5. Liberal-leaning articles tend to remain longer on the NYT homepage across several sections, with the notable exception of the "Style" section. This effect is particularly pronounced in the "Arts" section. In contrast, for the WSJ, an article's political slant generally does not have a statistically significant association with its homepage survival duration. One exception is the "US" section, where articles with a pro-Democrat slant tend to exit the homepage.[A4]

We can also focus on sections which are relatively unique to each newspaper, providing an opportunity to to test the prediction of supply-driven bias models that reduced competition increases bias. In the WSJ's "C Suite" section, liberal-leaning articles are far more likely to exit the homepage in any period (though the parameter is not statistically significant). This supports the supply-side bias model where editorial inclinations might have more influence in specialized content areas with less competition. Conversely, in the NYT's "NY Region" section, evidence for pronounced bias is weaker, although the limited sample size for this section might affect the findings.

[A4] Liberal-leaning articles in the "World", "Arts", and "Style" section tend to stay longer on the homepage. However, the parameters are not statistically significant.

Table A4: Survival Analysis of Homepage by Sections (Hazard Ratios)

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

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

p-stats are for hazard ratio differences from one.

Exponentiated coefficients; Standard errors in parentheses

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

The role of article sentiment also varies between the two newspapers (though the parameters are not significant). For the WSJ, sentiment has little impact on article duration on the homepage, except in the "C Suite" section where more positive articles tend to stay longer. In contrast, sentiment plays a more significant role in the NYT, particularly in the news section. Generally, articles with a more positive tone have longer survival times on the homepage, except in the US news section where more negative news tends to stay longer.

Table A5 presents homepage survival estimates for subsets of articles the newspapers tag with keywords corresponding to five major political topics. The topics are: "Politics", including news about political parties and elections; "Coronavirus", covering the pandemic and vaccinations; "Race/Gender", addressing issues such as discrimination and significant events like the George Floyd incident; "Abortion", involving news related to abortion, including Supreme Court decisions; and "Immigration", encompassing news on

immigration issues.

Table A5: 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***	1.079***	1.102***	0.981	0.975	0.971***	0.878***	0.905***	0.749**	0.948
	(0.00665)	(0.0135)	(0.0145)	(0.0288)	(0.0248)	(0.00967)	(0.0327)	(0.0305)	(0.0894)	(0.142)
Sentiment Scores	1.012	1.029	0.724***	1.106	0.865**	0.963	0.889	0.725***	1.195	0.856
(Abstract)	(0.0210)	(0.0381)	(0.0257)	(0.104)	(0.0584)	(0.0220)	(0.0974)	(0.0582)	(0.334)	(0.275)
Pro-Dem. Scores	0.862***	0.887*	1.107	0.621**	0.577***	1.216***	0.997	1.749***	1.139	1.061
(Abstract)	(0.0347)	(0.0627)	(0.0877)	(0.123)	(0.0871)	(0.0573)	(0.189)	(0.265)	(0.494)	(0.695)
# of New	1.023***	1.023***	1.017***	1.015	1.015*	1.038***	1.044***	1.048***	1.070***	1.039
Articles	(0.00240)	(0.00372)	(0.00424)	(0.0102)	(0.00791)	(0.00244)	(0.0106)	(0.00830)	(0.0247)	(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

p-stats are for hazard ratio differences from one.

Exponentiated coefficients; Standard errors in parentheses

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

Consistent with our prior findings, audience engagement, as indicated by tweet counts, is a key factor for longer homepage duration across most topic tags and for both newspapers. The exceptions are articles tagged with "Coronavirus" and "Race/Gender" in the NYT.

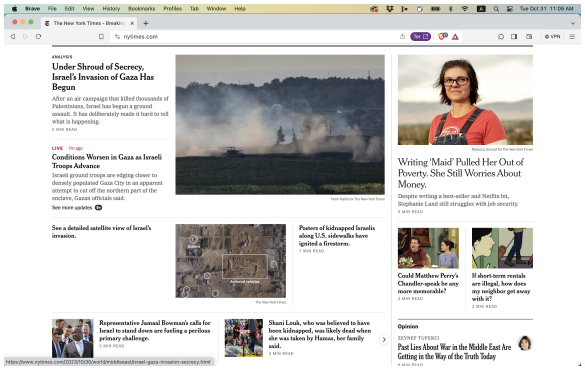
In terms of supply-side factors, in both papers there is a significant relationship between political slant and article homepage survival. The 'Politics" category is representative: more liberal articles in the NYT and more conservative articles in the WSJ tend to stay longer on the homepage. A one standard deviation increase in the pro-Democrat score results in a 2.4% decrease in the hazard rate for NYT articles and a 4% increase for WSJ articles. Similar patterns are observed in other topic tags, particularly "Abortion".

Appendix F Survival Analysis on Leading News

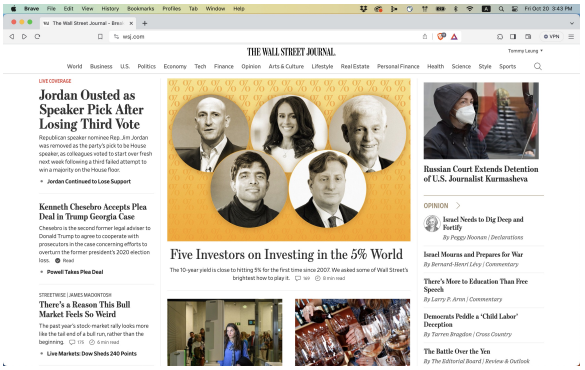
While our prior analysis focused broadly on article dynamics across the entire homepages of major newspapers, this section narrows the focus to the leading news—articles occupying the most prominent slots at the top of the homepage. These articles often set the day's narrative and attract the most immediate attention from visitors.[A5] Figure A4 illustrates the positions of leading news in the NYT and WSJ, respectively.

Figure A4: Leading News on NYT and WSJ Homepage

NYT



WSJ



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.[A6] 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 required a different approach due to the lack of position-specific markers in their URLs. We used the order in which articles are scraped from the API, where the first ten URLs in the sequence are typically those with the most

[A5] These articles get more attention than others on the homepage. For example 5.1% of the NYT's ten most viewed articles each hour are from the top of the homepage while the far higher number of articles on the remainder of the homepage make up 3.0% (the values are 7.2% / 2.8% for top ten shares and 2.8% / 2.3% for top ten emails).

[A6] The 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".

prominent positions on the NYT homepage. These were classified as leading news for the NYT.[A7]

It is important to note that the bulk of these leading articles are from 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 findings. Table A6 presents separate survival (at the leading news position of the homepages) estimates for each news section.

Table A6: Survival Analysis of Leading News by Major News Sections

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

p-stats are for hazard ratio differences from one.

Exponentiated coefficients; Standard errors in parentheses

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

Consistent with our previous findings, political slant significantly influences the editorial decisions for leading news selection, even after accounting for demand-side dynamics. In the NYT, liberal-leaning US and World news articles tend to remain in top homepage positions longer. A one-standard-deviation increase in pro-Democrat scores decreases the hazard rates by 3% for US news and 7% for World news, indicating a lower likelihood of swift replacement.

Conversely, the WSJ shows a distinct trend in its World news section, where conservative-leaning articles are more resilient in top positions. A one-standard-deviation rise in pro-Democrat scores corresponds to a 12% increase in hazard rates, suggesting that liberal-leaning World news articles in the WSJ are more likely to be quickly replaced.

Table A7 splits the samples by major news tags (these keywords are described in more detail in Appendix E). The results show that on the NYT homepage, articles tagged with

[A7] The analysis excludes Opinion, which has a fixed position below leading news in the NYT and is not scraped for the WSJ.

"Politics", "Coronavirus", "Abortion", or "Immigration" and have a liberal slant tend to dominate the top spots for extended periods. Conversely, in the WSJ, leading news articles tagged with "Politics" or "Race/Gender" and have a conservative orientation have a longer duration on their prominent homepage positions.

Table A7: 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***	1.079***	1.102***	0.981	0.975	0.971***	0.878***	0.905***	0.749**	0.948
	(0.00665)	(0.0135)	(0.0145)	(0.0288)	(0.0248)	(0.00967)	(0.0327)	(0.0305)	(0.0894)	(0.142)
Sentiment Scores	1.012	1.029	0.724***	1.106	0.865**	0.963	0.889	0.725***	1.195	0.856
(Abstract)	(0.0210)	(0.0381)	(0.0257)	(0.104)	(0.0584)	(0.0220)	(0.0974)	(0.0582)	(0.334)	(0.275)
Pro-Dem. Scores	0.862***	0.887*	1.107	0.621**	0.577***	1.216***	0.997	1.749***	1.139	1.061
(Abstract)	(0.0347)	(0.0627)	(0.0877)	(0.123)	(0.0871)	(0.0573)	(0.189)	(0.265)	(0.494)	(0.695)
# of New	1.023***	1.023***	1.017***	1.015	1.015*	1.038***	1.044***	1.048***	1.070***	1.039
Articles	(0.00240)	(0.00372)	(0.00424)	(0.0102)	(0.00791)	(0.00244)	(0.0106)	(0.00830)	(0.0247)	(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

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

Appendix G Print and Podcasts: Additional Analysis

In this section we present additional estimates of the print edition and podcasts tests of the supply-side model of bias (Section 6 of the text).

We begin with the print estimates for each news tone and for the overall sample. Table A8 examines whether an article is included in the print edition. Using a linear regression model with the binary outcome of print inclusion as the dependent variable, we find that articles featured on the homepage or those with a high number of tweets are more likely to be selected for the print edition. The article's pro-Democrat scores generally have a positive effect on inclusion in the print edition, though these effects are not statistically significant.

Table A8: NYT Print Edition Linear Regression by News Tone

	(1) News	(2) Opinion	(3) Feature	(4) All
Ever on Homepage	0.364***	0.190***	0.0845***	0.251***
	(0.00534)	(0.0322)	(0.00691)	(0.00421)
ln Tweets Count	0.0556***	0.0764***	0.0832***	0.0685***
	(0.00158)	(0.00489)	(0.00289)	(0.00134)
Sentiment Scores (Abstract)	0.0442***	-0.0385**	0.00304	0.0186***
	(0.00489)	(0.0154)	(0.00757)	(0.00402)
Pro-Dem. Scores (Abstract)	0.00499	0.0224	-0.0184	0.00439
	(0.00972)	(0.0284)	(0.0156)	(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

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

Next we turn to podcasts. The main text shows that for news tones there is a weak correlation between an article's political slant and its selection for a podcast. Table A9 presents estimates when divide the sample by new sections.[A8]

[A8] Certain sections, namely "Style" for the NYT and "Arts" for the WSJ, are entirely absent from their corresponding podcast features.

Table A9: Podcast Linear Regression by Major News Sections

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

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

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

These estimates are consistent with those from the main text: an article's political bias does not significantly influence its likelihood of being featured in a podcast. There are two exceptions, articles from the NYT's "Life" section and the WSJ's "Style" section deviate from this pattern. Interestingly, it is the conservative articles from the NYT's "Life" section and the liberal articles from the WSJ's "Style" section that are more likely to be featured in their respective podcasts. This unexpected result could be due to newspapers using podcasts to appeal to a more diverse audience, potentially reaching beyond their traditional political bases.