Journalist Ideology and the Production of News: Evidence from Movers

Levi Boxell*

Jacob Conway, *Stanford University*

December 2022

Abstract

What role do journalists play in determining the political slant of the news they produce? We develop a model where journalists and newspaper outlets contract over both slant and wages. The model implies a set of conditions under which we can consistently estimate the role of journalist preferences in driving the observed variation in slant across outlets by leveraging journalist transitions between outlets. To measure slant, we train a transformer-based, machine learning model using articles tweeted by politicians and apply it to a full-text database of 20+ million newspaper articles published in the US between 2013 and 2018. Applying our model-informed estimators to the data, our estimates (a) reject the hypothesis that journalists have zero ideological preferences over the content they produce and (b) imply that 16 percent of observed variation in slant across outlets can be explained by journalist preferences.

*E-mail: leviboxell@gmail.com, jcconway@stanford.edu. We thank John Dalton, Pascaline Dupas, Matthew Gentzkow, Andrew Hall, Tommy Leung, Jesse Shapiro, and Isaac Sorkin, as well as attendees at various presentations, including seminars at Stanford University and the IHS Summer Research Fellowship, for their feedback and suggestions. We also thank Brandon Williams for his excellent support with using ProQuest TDM Studio. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1656518. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also acknowledge funding from the Institute for Humane Studies.

1 Introduction

In the United States, 28 percent of journalists identify as a Democrat whereas only 7 percent identify as a Republican (Weaver et al. 2019).[1] How does the skewed distribution of political affiliation in the news industry affect the type of news that gets produced?

Some commentators have linked the labor market dynamics directly to the political slant of news as a whole, with Matthew Yglesias, co-founder of Vox, stating "*I think content is getting dragged left because that's the labor market (and to an extent the audience) rather than as deliberate strategy*" (Smith 2021).[2] High-profile exits of certain journalists are also linked to within-outlet political representation. For example, in 2017, *The New York Times* hired Bari Weiss to diversify the paper's opinion page. Yet, in 2020, Weiss resigned—citing "*constant bullying by colleagues who disagree with my views*" (Weiss 2020). Moreover, prior academic work has provided theoretical motivations for how journalist preferences could drive persistent media bias (e.g., Baron 2006).

In this paper, we empirically examine the extent to which individual journalists shape the ideological content of the articles they write and the market provision of slant as a whole.

We start by building a simple model of the journalist labor market and the production of news. Our model highlights how observed slant is a weighted combination of three components: journalist preferences, outlet average preferences, and a search-specific outlet shock. Our model provides testable predictions for whether journalists have preferences over slant, and it provides guidance for estimating the magnitude of these preferences given potential selection in job transitions induced by search-specific shocks to outlet preferences.

In order to take the model-informed estimators to the data, we construct a novel measure of media bias leveraging recent advances in machine learning. Using a new dataset of articles tweeted by politicians, we fine-tune RoBERTa (a transformer-based model of human language) to predict

[1]In 1971, 25.7 percent of surveyed journalists in the US identified with the Republican party and 35.5 percent identified with the Democratic party. By 2013, only 7.1 percent of surveyed journalists identified with the Republican party whereas 28.1 percent continued to identify with the Democratic party (Weaver et al. 2019). When focusing on Washington correspondents, the partisan skew is more striking—with 89 percent of surveyed correspondents indicating they voted for Clinton in the 1992 election. The distribution of self-reported partisan identification among journalists is mirrored by the overall distribution of political contributions from members of the news and media industry (Bonica 2013). See also Figure 1.

Academics and media critics have also expressed concerns about an overarching liberal media bias in the US—with these concerns being expressed in a diverse set of outlets including *The Washington Post*, *The Wall Street Journal*, and *The Quarterly Journal of Economics* (e.g., Groseclose and Milyo 2005; Wemple 2017; Sauter 2020).

[2]See also Groseclose (2011).

the political party of the politician who shared a given article. Our fine-tuned model performs well on both in-sample cross-validation tests and out-of-sample comparisons to human coders, and it represents a significant improvement over bag-of-word methods used previously in the literature.

We then apply our measure of slant to predict the partisan slant of more than 20 million articles published between 2013 and 2018 from over 300 newspapers in the ProQuest US Newsstream database (accessed via TDM Studio). Importantly, the ProQuest database also includes the byline for each article—allowing us to construct an author-level panel dataset that tracks the ideological slant of the articles written by a given journalist as they move across outlets.

Our empirical strategy for the observational data builds on the rich literature in labor economics that estimates worker and firm wage fixed effects (e.g., Abowd et al. 1999; AKM) or worker preferences (e.g., Sorkin 2018) using employee transitions between firms.

We start by estimating an event-study specification using a sample of consistent movers.[3] Under our model and assumptions, if we are able to reject the hypothesis that journalist close the entire origin-destination gap between outlets, we can also reject the hypothesis that journalists have zero preferences over the slant they produce. In other words, if journalists simply write content that mirrors the preferred slant of an outlet and have zero ideological preferences, the gap in slant between the origin and destination outlets would close entirely when journalists switch outlets. On the other hand, if journalists have strong ideological preferences, we would expect there to be little-to-no change in observed slant when a journalist switches outlets.

Our event-study estimates reject the hypothesis that journalists have zero ideological preferences over the slant they produce. Using our sample of consistent movers, journalists close roughly 75 percent of the gap in average slant between the origin outlet and the destination outlet. The 95 percent confidence intervals on our event study estimates exclude both 0 percent and 100 percent.

We also use the AKM fixed effect framework to decompose the variation in article-level slant using journalist and outlet fixed effects. In our preferred specification, restricted to the set of movers, we estimate that journalist fixed effects explain 20 percent of the observed variation in slant, outlet fixed effects explain 24 percent, and sorting between journalists and outlets explains another 11 percent.[4]

We also examine the relationship between journalist fixed effects and other observable char-

[3] Among journalists, there are a wide variety of work arrangements including both freelancers and salaried employees. In our event study analysis, we constrain our analysis to likely salaried employees by focusing on journalists who wrote T or more articles at the origin outlet before switching and writing T or more articles at the destination outlet.

[4] The remaining variation is explained by time fixed effects, other covariance terms, and the residual.

acteristics of journalists. As we would expect, the distribution of journalist fixed effects for Republican journalists is further to the right (i.e., more conservative) than the distribution for other journalists. And, consistent with the heterogeneity in underlying political preferences, women and non-white journalists tend to have more liberal fixed effect estimates than men and white journalists.

While interesting, we view these results as primarily *descriptive* given the search-specific shocks to outlet preferences highlighted in our model which would likely violate the conditional mean independence assumption of the AKM model.

We derive a new estimator to provide *causal* estimates of the *between* outlet variation in slant driven by journalist ideological preferences. By focusing on only the *between* outlet variation in slant, we are able to provide causal estimates under a weaker set of assumptions than the conditional mean independence assumption of the AKM model.[5] Rather than assuming no selection on the search-specific shocks (as the AKM model assumes), our estimator requires an "equal selection" assumption—that is, the conditional expectations of the search-specific shocks are the same for entering, exiting, and non-moving journalists at a given outlet.

Our preferred estimates from this specification suggest journalists explain 16 percent of the observed variation in average slant between outlets. This can be compared to the 20 percent of outlet-level variation in slant attributed to consumers in Gentzkow and Shapiro (2010).

Overall, these results suggest journalists' ideological preferences play an important role in driving the market provision of slant.

This paper relates to several strands of literature.

First, our paper relates to a literature on media bias. Within this literature, our paper makes both a methodological and a substantive contribution.

Methodologically, previous work has used bag-of-words or citation based measures of media bias (e.g., Gentzkow and Shapiro 2010; Groseclose and Milyo 2005). We leverage recent advances in machine learning, BERT-based transformer models, to measure media bias and show how these more advance methodologies can show substantial improvement in predicting the human perceived slant of an article. This builds on separate work using recent advances in machine learning to measure non-verbal media bias based on the facial expression of politicians in images used on

[5]Moreover, we may be particularly interested in the role of journalists in driving the *between* outlet variation in slant, rather than both the *between* and *within* outlet variation. For example, if consumers are choosing between outlets to subscribe to, between outlet variation in slant may be more relevant for understanding how variation in slant affects consumer welfare.

news websites (Boxell 2021).

We also make substantive contributions to the literature on media bias. Much of the prior empirical work on media slant has focused on demand-driven motives for media bias (e.g., Gentzkow and Shapiro 2010; Martin and Yurukoglu 2017), conglomeration and owner effects (e.g., Martin and Mccrain 2019), or professional norms (e.g, Baum and Groeling 2008; Shapiro 2016).

The two closest related papers empirically are Cagé et al. (2022) and Xu (2022).[6] Cagé et al. (2022) use an AKM-style decomposition to show that much of the variation in slant on French television and radio shows can be explained by host- or journalist-specific fixed effects. Xu (2022) also leverages journalist turnover to examines how financial journalists' social networks influence the favorability of coverage towards firms in the US.[7]

Our model, however, highlight the limits of the AKM-style variance decomposition estimates given the potential for selection into journalist transitions. We solve this selection issue by providing a novel estimator and associated set of conditions with which we can recover the role of journalist preferences in driving observed variation in slant across outlets. Overall, our empirical results corroborate Cagé et al. (2022) and Xu (2022) in finding important effects of journalists, but in the broader US print newspaper market as opposed to French broadcasting or US financial news.

Second, our paper relates to a literature on employee preferences and compensating differentials. Some work examines the importance of compensating differentials in general (e.g., Sorkin 2018), whereas other work focuses on more specific dimensions such as a firm's corporate social responsibility behavior (e.g., Hedblom et al. 2019) or the co-partisanship of co-workers (e.g., Gift and Gift 2015; Colonnelli et al. 2021). Our work provides new evidence on employee preferences over the ideological content of their production in the news industry.

2 Model of News Production and Empirical Implications

We build a simple model of news production with heterogeneous journalists and outlets.

[6]Baron (2006) showed theoretically how journalist discretion and preferences could drive persistent media bias.

[7]Barbera and Sood (2018) find a strong correlation between journalist campaign contributions and the ideological slant of the content they produce. See also Hassell et al. (2020) and Graves et al. (2016).

2.1 Journalists

Journalist i has utility

$$u_{it}(x,\omega) = \omega - \beta_i(x-\delta_i)^2 + \varepsilon_{it}$$

where x is the average slant of articles written, δ_i is the bliss point for journalist i, ω is an unobserved utility transfer from the outlet to the journalist (e.g., wages and benefits), and ε_{it} is a type-1 EV shock that is independent across individuals, outlet type, and time. Note that for a given match, utility can be transferred by giving journalists assignments closer to their bliss point δ_i or increasing the unobserved utility transfer (e.g., wages). We let ω_o denote the outside option for a journalist.

2.2 Outlets

An outlet j has preferences over their portfolio of journalists A_j

$$u_j(A_j) = R_j(A_j) - \sum_{i \in A_j} \omega_{ji}$$

where

$$R_j(A_j) = \sum_{i \in A_j} \left[\nu_j - \beta_j(\bar{x}_i - \delta_j - \xi_{ij})^2\right].$$

Here, ν_j is a fixed revenue per journalist, δ_j is an outlet's average slant ideal point, β_j is the strength of an outlet's preferences over slant, and ξ_{ij} is search-specific preference shock which is mean zero and independent across *searches*.

Outlets have preferences over revenue $R_j(A_j)$ minus the wage transfers to journalists $\sum_{i \in A_j} \omega_{ij}$.

2.3 Matching and Contracting

There is an initial distribution of journalists across outlets. Each period, firms draw a vector of search-specific preference shocks ξ_j of length s_j. Journalists are then matched to firm-searches, where journalist i is matched to firm j's s search with some probability $p_{ijs}(\cdot)$. We do not take a stand on the inputs to $p_{ijs}(\cdot)$—it could include previous values of slant x, previous wages ω, previous firms j, and/or previous or current search-specific shocks ξ.

The firm observes the journalist's current contract (x_0, ω_0) and type (δ_i, β_i) and gives a job offer $(\omega, \bar{x})$ that is composed of a utility transfer and an average slant. Journalists then observe their T1EV draw ε_{ijt} and maximize their utility by choosing among (a) their current position at their current firm j, (b) the position at the new firm j', and (c) their outside option. We assume journalist's output is perfectly observable to the firm, contract enforcement is costless, and both journalists and outlets are myopic.

Given a match with journalist i who had previous contract (x_0, ω_0), the new firm j' will choose a slant offer that solves

$$\max_{x,\omega} \underbrace{\Delta U_{j'}(x,\omega)}_{\text{Change in Profit}} \times \underbrace{M_i(x,x_0,\omega,\omega_0)}_{\text{Prob. Journalist Accepts Slant Offer}},$$

where

$$M_i(x,x_0,\omega,\omega_0) = \frac{\exp\left(\omega - \beta_i(x-\delta_i)^2\right)}{\exp(\omega_o) + \exp\left(\omega - \beta_i(x-\delta_i)^2\right) + \exp\left(\omega_0 - \beta_i(x_0-\delta_i)^2\right)}.$$

and

$$\Delta U_{j'}(x,\omega) = R_{j'}(A_{j'} \cup \{x\}) - R_{j'}(A_{j'}) - \omega.$$

Proposition 1. *The optimal slant contract is* $x^*_{ij} = \frac{\beta_i\delta_i + \beta_j(\delta_j + \xi_{ij})}{\beta_i + \beta_j}$, *and the optimal wage is* $\omega^*_{ij} = h(\xi_{ij}, x_0, \omega_0|\theta)$ *for some function h.*

Proof: *See Appendix A.*

Note that the optimal wage function $h(\cdot)$ is a function of the journalist's previous contract (x_0, ω_0), whereas the optimal slant contract x^*_{ij} is not.

It is important to note that while ξ_j is mean zero across *searches*, it is not necessarily mean zero over the set of *accepted* matches. For example, if journalists have strong slant preferences and if journalists are all more conservative than the most conservative outlet, then journalists will be more likely to accept a slant contract when $\xi_{ij} > 0$. And so, $\mathbb{E}(\xi_{ij}|i \in A(j)) > 0$. In this sense, *realized* variation in ξ can also be a function of journalist preferences.

A sufficient, but not necessary, condition for $\mathbb{E}(\xi_{ij}|i \in A(j)) = 0$ is if journalists do not have slant preferences.

Assumption 1. *(Wage Independent Search Shocks) The outside option* ω_o *and initial wage* ω_0 *of a journalist i matched with an outlet j is independent of the search specific shock* ξ_{ij}. *That is,* $\mathbb{E}(\omega_0|\xi_{ij}) = a$ *and* $\mathbb{E}(\omega_o|\xi_{ij}) = b$ *for* $a, b \in \mathbb{R}$.

Proposition 2. *If $\beta_i = 0$ for all i and the wage independent search shocks assumption holds, then $\mathbb{E}(\xi_{ij}|i \in A(j)) = 0$ and $\mathbb{E}(x^*_{ij}|i \in A(j)) = \delta_j$.*

Proof*: See Appendix A.*

It is important to note our proof of this proposition does not require us to take a stand on the inputs to the matching function besides requiring independence between the search-specific shock ξ and a journalist's outside option ω_o and initial wage ω_0. This still allows initial wages ω_0 and outside options ω_o to enter the search function—but they cannot vary with respect to ξ_{ij}. Moreover, this assumption still allows for interactions between initial slant x_0 and the search specific shock ξ in the search function, as long as wages and outside options are still independent.

3 Data

Our data is derived from three primary sources: articles shared by politicians on Twitter, full-text newspaper articles from ProQuest, and voter registration data from L2. We use the politician-shared articles to build a labeled training dataset, which allows us to train a machine learning model for predicting an article's political slant. We use the full-text data on newspaper articles to track journalists as they move across outlets and to measure how the slant of their content varies as they move. And we use the voter registration data to determine the ground-truth partisan affiliation of journalists in our data set.

3.1 Articles Tweeted by Politicians

We construct a labeled dataset of news articles shared by either Republican or Democratic politicians. The labeled nature of this dataset (either Republican or Democratic) is important for training our measure of media slant later.

We start with a list of active Twitter accounts by politicians from ProPublica's Politwoops dataset released in July 2019. We then scraped the tweets from each username listed—giving us the tweet history from more than 1800 usernames associated with politicians.

From the dataset of politician tweets, we locate all URLs tweeted by politicians—giving more than 1 million tweets with a URL.[8] We then randomly sample 600,000 URLs and scrape the text associated with each URL. Not all URLs are successfully scraped. We obtain 348,821 URLs with

[8]We exclude urls including "youtu.be," "pscp.tv," "facebook.com," "instagram.com," "patreon.com," and "google.com."

some text. We further exclude URLs with either (a) 1,000 or less characters or (b) no author detected by our scraper. This gives a final dataset of 193,883 non-unique URLs with text. 59,363 of these URLs were shared by Republican politicians, and 134,520 of these URLs were shared by Democratic politicians.

3.2 ProQuest Full-Text News

We also use a full-text database of news articles from more than 300 US newspapers, including major US daily newspapers such as *The New York Times* and *The Wall Street Journal*. The full-text nature of the data enables us to improve upon bag-of-word or keyword-based approaches to measuring media slant (e.g., Gentzkow and Shapiro 2010) by applying recent machine learning advances that leverage textual context (e.g., BERT or RoBERTa), thereby allowing us to construct more accurate measures of slant at the article level. Moreover, the large number of outlets allows us to track journalists as they move across outlets and provides insights across both national and local newspaper markets.

We obtain the full text for all articles from the ProQuest US Newsstream (accessed via TDM Studio) published between January 2013 and December 2018. This provides more than 20 million articles. For most of the analysis, we restrict to articles labeled as "News" by ProQuest.

For articles with multiple authors on the same byline, we reshape the data to be at the author-article level. So, an article that is written by two authors would appear twice in our dataset—once for each author.[9] (From here-on, when we refer to an "article," we mean an "article-author" pair.) We restrict attention to authors with more than 10 articles in our data and perform various cleaning steps in an attempt to remove wire services, syndicated articles, and outlet mergers.[10]

After the restrictions, we are left with a dataset of 7,873,626 articles from 31,985 authors published in 305 outlets. Table 1 provides summary statistics for our sample. Roughly a third of the authors in our dataset write for the same outlet throughout our analysis period, while the remaining two-thirds of authors switch outlets at some point. However, there appears to be heterogeneity in the share of movers versus non-movers across outlets. For example, many authors at *The Wall*

[9] We assume any individual with the same name is the same person, and we clean author names by replacing common female and male nicknames using the list of nicknames from https://github.com/uguryi/abeR/tree/master/inst/extdata, which is derived from Abramitzky et al. (2021).

[10] We drop observations that are non-unique across author, slant, year, and month. We also drop observations for authors in any month in which the author appears in a wire service (TCA Regional News, University Wire, CNN Wire Service, CNN Spanish Wire, TCA News Service, Gannett News Service, US Newswire, or any other source that includes "wire" in its name). We combine evening and day editions, city-specific editions, online editions, and outlet mergers into a single outlet where known.

Street Journal never write for another outlet, whereas the vast majority of authors at *The Chicago Tribune* switch outlets at some point. These differences may be a combination of differences in the use of freelancers versus staff writers, job stability, syndication agreements, and the use of wire services.

3.3 Voter Registration Data

We also use voter registration data from L2 to match journalists to their party registration status. To match journalists, we first apply a similar name cleaning procedure to the L2 data. We then assign a journalist a specific party if matched on first and last name to a set of voters with a unique *party,* after restricting to voters in the same state as the outlet headquarters. This allows a journalist to be matched to multiple voters in a given state, as long as those voters all have the same party affiliation. The party assignment is done at the author-outlet level—thus allowing authors to change party affiliation as they change outlets.

Of the more than 8,000 journalists we matched to voter registration records, 49 percent registered as Democrats, 27 percent as non-partisan, and 21 percent as Republican.

4 Measuring Political Slant and Stylized Facts

We use the politician-shared articles as a labeled dataset to fine-tune an existing transformer-based machine learning model (RoBERTa) to predict the political party of the politician sharing each article based on the article's text. Our fine-tuned RoBERTa model performs well in the leave-out validation set and is a significant improvement over previous methods when evaluated on a separate human-labeled dataset of political news articles.

4.1 RoBERTa Model

To measure political slant, we use the RoBERTa model framework (Liu et al. 2019). The RoBERTa model builds on the previously successful BERT transformer architecture (Devlin et al. 2019). (See Vaswani et al. 2017 for more on transformers.)

BERT stands for Bidirectional Encoder Representations from Transformers and is an unsupervised pre-trained model. To start, BERT uses word embeddings for each token from a 30,000 token vocabulary taken from Wu et al. (2016). These word embeddings are then split into sequences at the sentence level (each sequence having two sentences). The BERT model is the trained under a dual objective function. First, the BERT model will randomly "mask" or hide certain tokens (i.e., words) and the training objective is to successfully predict the masked tokens based on the surrounding context. Second, the BERT model is tasked with predicting the next sentence in a sequence. A key contribution of BERT is the ability to use the left *and* right context simultaneously in making these masked word predictions, as opposed to the unidirectional or left-to-right predictions used in many other language models (e.g., OpenAI's GPT). The BERT model was pre-trained on data from the BooksCorpus (Zhu et al. 2015) and English Wikipedia, giving a combined total of 3.3 billion words.

After pre-training the BERT model on the masked word task, the model can be subsequently fine-tuned on various downstream tasks, such as next sentence prediction, classification, or question and answering. When released in 2018, BERT achieved state-of-the-art performance on many of the standard NLP tasks including both sentence-level and token-level tasks. The base BERT model has 12 layers, a hidden size of 768, and 12 self-attention heads—giving a model with 110 million parameters. The large version of the BERT model has 340 million parameters.

RoBERTa, or robustly optimized BERT approach, is one of several refinements to the BERT model. It improves on BERT's performance by pre-training on more data and larger batches, training on longer sequences, dynamically changing the masking pattern, and removing the next sentence prediction objective. These modifications resulted in significant improvements over BERT across multiple NLP tasks.

4.2 Fine-Tuning RoBERTa and Measurement of Slant

The RoBERTa model is a pre-trained model for masked word prediction. We use the base RoBERTa model available in the HuggingFace library (https://huggingface.co/roberta-base).

Our objective is to accurately predict the *perceived* political slant of an article. That is, how liberal or conservative would a reader think a given article is? To achieve this objective, we use our dataset of articles tweeted by politicians to fine-tune the RoBERTa model.

To ensure our training dataset is balanced, we randomly select 57,000 articles tweeted by Republican politicians and 57,000 articles tweeted by Democratic politicians. We label each article according to the party of the politician who tweeted it. We then randomly select 70 percent of the articles to be the training dataset and the remaining articles to be the validation dataset.[11]

[11] Since an article can be tweeted by more than one politician and we allow duplicate articles in our dataset in general, we exclude all articles in the validation dataset that match the text of an article in the training dataset. This helps

Since the RoBERTa architecture assumes a sequence length of 512 tokens (where one token is roughly one word), we restrict the initial sections of each article when they are longer than the required sequence length and we "pad" articles that are shorter than required. Our fine-tuning task is a sequence classification task with two labels (Republican or Democratic). We use the default training parameters from the HuggingFace library for RoBertA (e.g., cross-entropy loss), except that we use: a batch size of 64, 20 percent attention dropout, 20 percent hidden dropout, a learning rate of 0.5×10^{-5}, a linear learning rate scheduler with 1000 step warmup, and 7 training epochs. These hyper-parameters were chosen by a manual search over values.

We choose the model weights that achieve the lowest cross-entropy loss in the validation dataset at the end of each epoch, which was achieved on the fourth epoch. The training loss was 0.286, the validation loss was 0.455, and the accuracy in the validation dataset was 0.809.

Using the fine-tuned RoBERTa model, we then predict the partisanship of every article in our ProQuest dataset.

4.3 External Validation of Slant Measure

In addition to the leave-out validation dataset of articles tweeted by politicians, we validate our measure of slant by comparing to human-coded measures of slant at the article level from Budak et al. (2016).

The Budak et al. (2016) dataset contains a sample of over 10,000 political news articles from 15 popular news outlets in 2013. Each article was then rated by MTurk coders on a 5-point likert scale for its favorability towards the Republican Party and, separately, for its favorability towards the Democratic Party. Each article was assigned to two coders, and human-perceived partisanship is defined as the relative difference in perceived favorability towards the Republican Party versus the Democratic Party.

For a random subset of these articles, we scraped the underlying text from the source URL and predicted the slant using three different methods (a) our RoBERTa slant measure described above, (b) a FastText model (Joulin et al. 2017) that we also train on the politician-tweeted articles, and (c) the Gentzkow et al. (2019) measure of partisanship from the 114th Congress.[12] Panel A of Figure 2 reports the Spearman rank correlation for each of the three text-based, machine-coded measures of slant and the human-coded slant measure from Budak et al. (2016). While all three

prevent cross-contamination of our training and test datasets.

[12] An alternative would be to retrain the Gentzkow et al. (2019) measure on the politician tweeted articles, which would likely improve the performance relative to using the bigrams from the 114th Congress.

machine-coded methods have a significant positive correlation with the human-coded measure, our RoBERTa-based slant measure performs significantly better than the other methods.

Panel B of Figure 2 reports results from a partitioning exercise where we examine the ability of each method to accurately partition the articles in half based on their relative partisan leanings. Our RoBERTa slant measure accurately categorizes over 70 percent of articles that are perceived to be clearly partisan by human coders—again, performing better than these other methods.

Overall, we believe these results suggest that both (a) our RoBERTa-based slant measure is performing well in capturing text-based variation that humans would perceive as political slant, and (b) our RoBERTa-based slant measure is a significant improvement over previous methods used in the literature.

Appendix B reports additional results comparing our model's performance against human-coded articles, as well as comparisons to previous measures of slant or partisanship in the literature.

4.4 Stylized Facts

4.4.1 Sorting of Journalists

In Figure 3, we compare the average slant at an outlet to the share of partisan journalists matched in the L2 data who are Republicans. Overall, there is a strong, positive, and significant correlation between the average slant of outlets and the share of Republican journalists.

The New York Times and *The Wall Street Journal* are interesting reference points. While both outlets are located in New York City, and thus can pull from the same potential set of employees, the share of Republicans at *The Wall Street Journal* is roughly twice that of *The New York Times*. And, correspondingly, the average slant at *The Wall Street Journal* is to the right of the average slant at *The New York Times.*

4.4.2 The New York Times *and* The Wall Street Journal

We examine the distribution of author-level slant at *The New York Times* and *The Wall Street Journal* across all authors and separately for the set of authors who have written for both outlets.

Panel A of Figure 4 plots the distribution of author-level slant at *The New York Times* and *The Wall Street Journal.* Consistent with prior suggestion, we see that the average slant of all authors at *The New York Times* is shifted to the left relative to the average slant of authors at *The Wall Street Journal*.

Panel B of Figure 4 is the same as Panel A, except it restricts attention to the set of authors who have written for both outlets. Among this set of authors, we see that the distributions of slant at *The New York Times* and *The Wall Street Journal* are more similar. Through the lens of our model, the set of authors who have written at both outlets either have significantly different ideal points δ_i than the general set of authors at *The Wall Street Journal* and/or have large search-specific shocks ξ.

Lastly, Appendix Figure B1 examines the average slant of select politicians when writing op-eds at *The New York Times* and *The Wall Street Journal.* As we would expect, among these politicians, the average slant of Republican politicians is to the right of the distribution at both outlets relative to the average slant of Democratic politicians.

4.4.3 Trends in Slant

We also examine trends between 2013 and 2018 in slant at key outlets, as well as the newspaper market overall. Figure 5 shows a secular shift to the left over time in slant for key outlets as well as the market as a whole. While this trend is gradual for the market as a whole, the 2016 election marks a turning point for many national news outlets. Several national US dailies exhibit a rapid shift leftwards after the 2016 election including for *The Wall Street Journal*—an outlet typically viewed as more favorable towards Republicans. Of the major national US dailies examined, it is interesting to note that while *USA Today* had similar levels of slant to *The Washington Post* in 2013, it reaches parity with *The Wall Street Journal* in the aftermath of the 2016 election—highlighting how outlets differentially responded to the election. *USA Today* has a higher share of Republican journalists than the other outlets considered (see Figure 3).

5 Preliminary Evidence from Journalist Movers

We now turn to examining the role of journalists in determining the slant of the content they produce.

We start by providing preliminary evidence on journalist content preferences using standard mover event study frameworks and AKM variance decompositions. We provide guidance on the extent to which we can consider these estimates causal, as opposed to descriptive, in light of our model.

To make progress empirically, we assume that journalists and outlets are only heterogenous in their bliss points δ and have common strength of preferences. That is, $\beta_i = \beta$ for all i and $\beta_j = 1$ for all j, where $\beta_j = 1$ is a normalization. Therefore,

$$x^*_{ij} = \frac{\beta}{\beta+1}\delta_i + \frac{1}{\beta+1}(\delta_j + \xi_{ij}).$$

Note that, in this case, the average observed slant at an outlet is also additively separable

$$\bar{x}_j = \frac{1}{|A(j)|}\sum_{i\in A(j)} x^*_{ij} = \frac{\beta}{\beta+1}\bar{\delta}_{A(j)} + \frac{1}{\beta+1}\delta_j + \frac{1}{\beta+1}\bar{\xi}_j$$

where $\bar{\delta}_{A(j)} = \frac{1}{|A(j)|}\sum_{i\in A(j)}\delta_i$, $\bar{\xi}_j = \frac{1}{|A(j)|}\sum_{i\in A(j)}\xi_{ij}$, and, recall, that $A(j)$ is the portfolio of journalists working at firm j.

5.1 Mover Event Studies

Consider a mover from j to j'. Then

$$x_{ij'} - x_{ij} = \frac{1}{\beta+1}\left[\left(\delta_{j'} + \xi_{ij'}\right) - \left(\delta_j + \xi_{ij}\right)\right].$$

Ideally, we would be able to estimate

$$\hat{x}_{it} = \gamma_i + d_{m(i,t)} + \sum_{l=-T, l\neq -1}^{l=T-1} \rho_l \left[\left(\delta_{j(i,t)} + \xi_{i,j(i,t)}\right) - \left(\delta_{o(i)} + \xi_{i,o(i)}\right)\right]\mathbf{1}_{\{t-T-1=l\}} + \varepsilon_{it},$$

where $\hat{x}_{it}$ is the estimated slant of the tth article written by journalist i, $t = 0$ is first article at the new outlet, $d_{m(i,t)}$ are calendar month fixed effects, γ_i are journalist fixed effects, ρ_l are the event study coefficients, and $j(i,t) \in \{o(i), d(i)\}$ is the outlet for journalist i in time t—either the origin or the destination.

The difference between ρ_l before and after the move would be an estimate of $\frac{1}{\beta+1}$. However, the outlet preference terms $\left(\delta_{j(i,t)} + \xi_{i,j(i,t)}\right)$ are unobserved. Instead, we estimate the following event-study specification

$$\hat{x}_{it} = \gamma_i + d_{m(i,t)} + \sum_{l=-T, l\neq -1}^{l=T-1} \rho_l (\bar{x}_{j(i,t)} - \bar{x}_{o(i)})\mathbf{1}_{\{t-T-1=l\}} + \varepsilon_{it},$$

replacing the outlet preferences with the observed average slant at the outlet $\bar{x}_{j(i,t)}$.

In general, $\mathbb{E}(\bar{x}_{j(i,t)} - \bar{x}_{o(i)}) \neq \delta_{j(i,t)} - \delta_{o(i)}$ and so our estimates $\hat{\rho}$ do not have precise mappings to the underlying model primitives. However, consider the case where $\beta = 0$. Under Proposition 2, we would have $\mathbb{E}(\bar{x}_{j(i,t)} - \bar{x}_{o(i)}) = \delta_{j(i,t)} - \delta_{o(i)}$ and $\mathbb{E}(\xi_{ik}|i \in A(j)) = 0$, and therefore, $\mathbb{E}(\hat{\rho}_1) = \frac{1}{\beta+1} = 1$. In this sense, by testing for whether $\hat{\rho}_{t\geq 0} = 1$, we are testing for whether journalists have preferences over the ideological content which they produce, i.e., $\beta > 0$. Thus, while this exercise does not allow us to pin down the relative magnitude of β, it is still informative about the existence of journalist preferences. An important caveat is that unobserved shocks ξ still function as measurement error in the independent variable, which could attenuate our estimates of ρ.

In estimating the event study specification, we restrict to "consistent" moves which we define as when a journalist has written T or more articles consecutively at the origin outlet before writing T or more outlets consecutively at the destination outlet.[13] Lastly, to improve precision of our estimates, we weight observations by the square of the relative move size between outlets $(\bar{x}_{d(i)} - \bar{x}_{o(i)})^2$. Figure 6 plots the distribution of move sizes $\bar{x}_{d(i)} - \bar{x}_{o(i)}$ for our estimating sample of consistent movers when $T = 5$. As we would like to see, the distribution of moves is approximately symmetric around zero and has many "large" moves.

Our event study estimates allow us to reject the hypothesis that $\rho_1 = 1$, and therefore, reject the hypothesis that $\beta = 0$. Figure 7 reports the estimates of ρ from our event study specification when $T = 5$ and when $T = 10$. The estimates show no significant pre-trends prior to the move. Upon move, both the $T = 5$ and $T = 10$ specifications show a quick transition—going from approximately zero to fluctuating around 0.75. The magnitude of this shift suggests journalists close roughly 75 percent of the observed slant gap between outlets when they move between outlets. We see no evidence of learning behaviors—with the estimates being relatively flat across time in the post-move window of analysis. If we were to interpret our event study coefficient estimates as estimates of $\frac{1}{\beta+1}$, they would imply $\beta \approx 1/3$ and so journalist preferences over slant would be a third of the size of outlet preferences over slant.

In Figure 8, Panel A repeats our estimation separately for the set of movers who move to a more conservative outlet (e.g., $\bar{x}_{d(i)} - \bar{x}_{o(i)} > 0$) and the set of movers who move to a more liberal outlet (e.g., $\bar{x}_{d(i)} - \bar{x}_{o(i)} < 0$). Panel B examines the relationship between the change in average outlet slant $\bar{x}_{d(i)} - \bar{x}_{o(i)}$ with the change in average journalist slant non-parametrically. Both panels suggest a high degree of symmetry across the direction and size of journalist moves across outlets.

[13] Prior to defining a "consistent" move, we drop observations of three or fewer articles at an outlet to reduce the potential influence of syndications and new wires.

The Appendix reports additional results.

5.2 AKM Decompositions

In this section, we continue to examine the role of journalists across a broader set of movers using a variance decomposition of journalist and outlet fixed effects following Abowd et al. (1999; AKM).

To highlight the appropriate interpretation of our AKM estimates, we can write and re-arrange the optimal slant contract as

$$x_{ij}^{*} = \underbrace{\frac{\beta}{\beta+1}\delta_i}_{\gamma_i} + \underbrace{\frac{1}{\beta+1}\delta_j}_{\alpha_{j(i)}} + \underbrace{\frac{1}{\beta+1}\xi_{ij}}_{\varepsilon_{ij}}$$

and so,

$$x_{ij}^{*} = \gamma_i + \alpha_{j(i)} + \varepsilon_{ij}.$$

As explained in Abowd et al. (1999), the outlet fixed effects α_j are identified by the set of journalists moving between outlets. In order for our estimates of the journalist and outlet fixed effects (γ_i and $\alpha_{j(i,t)}$) to be unbiased, we need AKM's conditional mean independence assumption to hold

$$\mathbb{E}(\varepsilon_{ij} | i \in A(j)) = 0.$$

If the conditional mean independence assumption holds, then we can use the leave-out estimator proposed by Kline et al. (2020) to consistently estimate the variance of the journalist fixed effects $V(\gamma_i)$, which is combination of the variation in journalist ideal points and their relative strength of preferences

$$V(\gamma_i) = \left(\frac{\beta}{\beta+1}\right)^2 V(\delta_i).$$

However, given the importance of selection in our model, we do not expect the conditional mean independence assumption to hold. As such, we interpret our AKM variance decompositions as a *descriptive* exercise familiar to many readers—not as estimates of underlying model primitives.

In practice, we estimate

$$\hat{x}_{ijt} = \gamma_i + \alpha_{j(i,t)} + \tau_t + e_{ijt}$$

to allow for systematic temporal variation in slant due to the timeliness of different topics, where

$\hat{x}_{ijt}$ is the average slant of articles published by journalist i at outlet j in month t, γ_i are journalist fixed effects, $\alpha_{j(i,t)}$ are outlet fixed effects, τ_t are time fixed effects, and e_{ijt} is the error term.

Relative to our event study sample, we are less restrictive in the set of articles we include. We still drop "short moves" of three or fewer articles at an outlet to reduce the influence of syndicated articles and news wires, but make no other restrictions after the restrictions defined in Section 3. This will include freelancers, salaried journalists, and other writers. We also collapse the data to the journalist-month level in our baseline specification to reduce noise in the outcome variable. In total, there are 30,048 journalists (with 5,513 of those being movers), 298 outlets, and 701,478 journalist-source-month observations in the connected sample.

Table 2 reports the variance decomposition estimates using both the plug-in estimator (Panel A) as well as the leave-out variance estimator proposed by Kline et al. (2020), which addresses the limited mobility bias that is often present in these employee-employer transition datasets (Panel B). Both panels restrict attention to the leave-out connected set of firms. The different columns report estimates on various samples of the data.

Panel B is our preferred set of estimates.

Our baseline specification (Column 1 of Panel B) estimates 30 percent of the variation in journalist-month slant is explained by journalist fixed effects, 23 percent is explained by outlet fixed effects, and 12 percent is explained by the covariance between journalist and outlet fixed effects. These results are reasonably robust when the estimation sample includes both movers and stayers (e.g., Columns 1, 4, and 5 of Panel B), with the exception of the specification where we replace outlet fixed effects with parent company fixed effects (Column 10 of Panel B) or restrict to "top" outlets (Column 6 of Panel B).

When our estimation sample is restricted to a set of movers only (Columns 2, 3, 7, 8, and 9 of Panel B), our estimates for the share of the variance explained by journalist fixed effects falls significantly and the share of the variance attributed to outlets increases. However, even in the most restrictive set of movers (Columns 8 and 9 of Panel B), 14 percent of the variation in slant is still explained by journalist fixed effects.

When restricting the sample to the set of movers *and* restricting to large moves (e.g., moves greater than 0.1), Column (2) still estimates a large role for journalist fixed effects. Column (8), which uses the $T = 5$ event study sample, reports smaller role for journalists fixed effects, but also reports a smaller role for outlet fixed effects—potentially driven by the limited sample size.

It is also interesting to note that both sorting and the variance of outlet fixed effects is larger among low-tier outlets. When restricting to the top 20 outlets by circulation, our estimated correlation between γ_i and α_j is 0.091 (Column 6, Panel B) compared to 0.222 in our full sample (Column 1, Panel B). These results could be consistent with a model where top-tier outlets differentiate themselves vertically, whereas low-tier outlets differentiate themselves horizontally.

Figure 9 plots the distribution of estimated journalist fixed effects for various subsets of journalists. Panel A reports the distribution for Republican, Democratic, and non-partisan journalists, where we see that the distribution of fixed effects for Republican journalists is significantly to the right of other journalists. Consistent with our understanding of underlying heterogeneity in political preferences across demographic groups, Panels B and C suggest that women have more left-leaning ideal points than men and non-white journalists have more left-leaning ideal points than white journalists.

6 Outlet-level Decompositions

Both the event study and AKM results are limited by their ability to address the selection driven by the search-specific shock process in our model. In this section, we relax the conditional mean independence assumption by focusing on estimating the role of journalists in driving the observed variation in *outlet-level* slant.

6.1 Model-Driven Specification

Recall that the average observed slant at an outlet is

$$\bar{x}_j = \frac{1}{|A(j)|}\sum_{i\in A(j)} x^*_{ij} = \frac{\beta}{\beta+1}\bar{\delta}_{A(j)} + \frac{1}{\beta+1}\delta_j + \frac{1}{\beta+1}\bar{\xi}_j.$$

Let

$$\psi_j = \frac{\beta}{\beta+1}\bar{\delta}_{A(j)}$$

denote the “journalist effect” at outlet j.

Our goal is to estimate $V(\psi_j)$.

Our identification strategy uses a movers identification strategy to difference out most of the non-ψ_j components, and we provide conditions under which we can ignore the remaining selection. Note that, for a mover i who moves from j to j', we can write

$$x_{ij'} - x_{ij} = \frac{1}{\beta+1}\left[\left(\delta_{j'} + \xi_{ij'}\right) - \left(\delta_j + \xi_{ij}\right)\right],$$

and we can write the difference in average slant between outlets j' and j as

$$\bar{x}_{j'} - \bar{x}_j = \frac{\beta}{\beta+1}(\bar{\delta}_{A(j')} - \bar{\delta}_{A(j)}) + \frac{1}{\beta+1}(\delta_{j'} - \delta_j) + \frac{1}{\beta+1}(\bar{\xi}_{j'} - \bar{\xi}_j).$$

Let $y_i = (\bar{x}_{j'} - \bar{x}_j) - (x_{ij'} - x_{ij})$. Combining the two equations above, we get

$$y_i = \frac{\beta}{\beta+1}(\bar{\delta}_{A(j')} - \bar{\delta}_{A(j)}) + \frac{1}{\beta+1}(\bar{\xi}_{j'} - \bar{\xi}_j) - \frac{1}{\beta+1}(\xi_{ij'} - \xi_{ij}),$$

which we can re-arrange and relabel as

$$y_i = \underbrace{\frac{\beta}{\beta+1}\bar{\delta}_{A(j')}}_{\psi^d_{j'}} - \underbrace{\frac{\beta}{\beta+1}\bar{\delta}_{A(j)}}_{\psi^o_j} - \underbrace{\frac{1}{\beta+1}\left[(\xi_{ij'} - \bar{\xi}_{j'}) - (\xi_{ij} - \bar{\xi}_j)\right]}_{e_i}$$

to get

$$y_i = \psi^d_{j'} - \psi^o_j - e_i. \tag{1}$$

Note that $\psi^d_j = \psi^o_j$ for a given outlet j.

Intuitively, $x_{ij'} - x_{ij}$ differences out journalist i's fixed preferences δ_i that remain the same across outlets, thereby leaving the difference in outlet components $(\delta_{j'} + \xi_{ij'}) - (\delta_j + \xi_{ij})$. Then, by subtracting $x_{ij'} - x_{ij}$ from $\bar{x}_{j'} - \bar{x}_j$, we difference out the outlet components that remain the same for all journalists (i.e., δ_j). Thus, leaving the difference in journalist components ψ_j between origin and destination outlets, as well as the difference in the structural errors for movers and non-movers at each outlet, $\left[(\xi_{ij'} - \bar{\xi}_{j'}) - (\xi_{ij} - \bar{\xi}_j)\right]$.

We would like to estimate equation (1) using origin and destination fixed effects among the sample of movers and to construct $\hat{V}(\hat{\psi}^d_j)$, an estimate of $V(\psi_j)$.

Two empirical issues remain.

First, it is well known that plug-in estimates of quadratic forms, such as variances, are biased (e.g., Kline et al. 2020). To address this, we use the covariance between the origin and destination fixed effects $cov(\hat{\psi}^d_j, \hat{\psi}^o_j)$ as our preferred estimator of $V(\psi_j)$.

Second, our estimates of ψ^d_j and ψ^o_j are potentially biased themselves if they are correlated with e_i—which is likely under any form of journalist preferences over slant or selection in job transitions. We postulate a set of conditions for which we can derive consistent estimates for $V(\psi_j)$ below.

Let T^o_{ij} be an indicator for whether journalist i's *origin* is outlet j, and let T^d_{ij} be an indicator for whether journalist i's *destination* is outlet j.

Assumption 2. *(Equal Selection) For all $j \in \mathscr{J}$,*

$$\mathbb{E}(\xi_{ij}|T^d_{ij}=1)=\mathbb{E}(\xi_{ij}|T^o_{ij}=1)=\bar{\xi}_j.$$

Remark 1. Note that the equal selection assumption is less restrictive than AKM's conditional mean independence assumption which would imply

$$\mathbb{E}(\xi_{ij}|T^d_{ij}=1)=\mathbb{E}(\xi_{ij}|T^o_{ij}=1)=\bar{\xi}_j=0$$

for all $j \in \mathscr{J}$.

Consider the linear model

$$\mathbb{E}(\mathbf{Y}|\mathbf{T})=\mathbf{T}\theta$$

where $\mathbf{T}=[-T^o_1,...,-T^o_J,T^d_1,...,T^d_{J-1}]$, a $N_m\times(2J-1)$ matrix and θ is a $(2J-1)$ parameter vector which we can partition as $\theta=[\theta^o,\theta^d]'$. Then

$$\hat{\theta}=(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{Y}$$

is the OLS estimator.

Proposition 3. *If the equal selection assumption holds, $\hat{cov}(\hat{\theta}^d,\hat{\theta}^o)$ is a consistent estimator for $\mathbb{V}(\psi_j)$.*

***Proof**: See Appendix.*

Remark 2. $\hat{\mathbb{V}}(\hat{\theta}^d)$ or $\hat{\mathbb{V}}(\hat{\theta}^o)$ would also be consistent estimators for $\mathbb{V}(\psi_j)$. Our preferred estimator is $\hat{cov}(\hat{\theta}^d,\hat{\theta}^o)$ because it reduces the finite sample or limited-mobility bias (see also Kline et al. 2020).

The equal selection assumption is admittedly strong. However, it is a weaker assumption than the conditional mean independence assumption commonly used in AKM and related models. Intuitively, while the conditional mean independence assumption says there is *no selection,* our working assumption allows for selection, but requires that it is the same across both sets of movers, as well as non-movers, for each outlet j.

Moreover, we can empirically test part of our the equal selection assumption. Under our assumptions and model, $\mathbb{E}(\hat{\theta}^d) = \mathbb{E}(\hat{\theta}^o)$. The extent to which these two estimates differ tells us something about the extent to which movers *to* outlet j are differentially selected compared to movers *from* outlet j—providing a partial test of the equal selection assumption.

Under certain conditions, our preferred estimator for $\mathbb{V}(\psi_j)$ is a lower bound for the variance in observed outlet-level slant driven by journalist effects *and* selection.

Assumption 3. *(Non-negative Covariance) The correlation between the average journalist ideal point and the average structural error at outlet j is non-negative. That is,* $cov(\bar{\delta}_{A(j)}, \bar{\xi}_j) \geq 0$*.*

Corollary 1. *If the equal selection and non-negative covariance assumptions hold, then* $\mathbb{V}(\psi_j + \frac{1}{\beta+1}\bar{\xi}_j) \geq \mathbb{V}(\psi_j)$ *and* $c\hat{o}v(\hat{\theta}^d, \hat{\theta}^o)$ *is a consistent lower bound for* $\mathbb{V}(\psi_j + \frac{1}{\beta+1}\bar{\xi}_j)$*.*

Proof*: Note that*

$$\mathbb{V}(\psi_j + \frac{1}{\beta+1}\bar{\xi}_j) = \mathbb{V}(\psi_j) + \mathbb{V}(\frac{1}{\beta+1}\bar{\xi}_j) + 2cov(\frac{\beta}{\beta+1}\bar{\delta}_{A(j)}, \frac{1}{\beta+1}\bar{\xi}_j) \geq \mathbb{V}(\psi_j),$$

with the last inequality holding due to the non-negative covariance assumption. Proposition 3 then implies $c\hat{o}v(\hat{\theta}^d, \hat{\theta}^o)$ *is a consistent estimator of a lower bound for* $\mathbb{V}(\psi_j + \frac{1}{\beta+1}\bar{\xi}_j)$*.*

6.2 Results

To visualize each of these differences in y_i, we partition the outlets into above- and below-median slant outlets using our full dataset and compute $\bar{x}_{j'}$ and $\bar{x}_j$. We then compute $x_{ij'}$ and x_{ij} for the set of movers who move from above-median outlets to below-median outlets, and vice-versa, from our $T = 10$ event study dataset as used in Section 5.1. (Note that this implicitly restricts to larger moves.)

Figure 10 plots the results. The "Outlet Gap" in each panel of Figure 10 reports $(\bar{x}_{j'} - \bar{x}_j)$, the difference in average slant between destination and origin outlet. The "Change on Move" in each panel reports $(x_{ij'} - x_{ij})$, the average change in slant for a journalist that moves from the origin to the destination outlet.

If outlets in the set of above-median outlets have homogenous preferences (and likewise for below-median outlets), then the difference between the "Outlet Gap" and the "Change on Move" for each panel would give us estimates of ψ^d and $-\psi^o$ for one set of outlets normalized relative to the other set. Using the below-median outlets as the reference group, our example implies $\hat{\psi}^d = (0.13 - 0.09) = 0.04$ and $\hat{\psi}^o = -(-0.13 + 0.11) = 0.02$ for the above-median set of outlets. And, thus, roughly a fifth of the difference in average slant between above-median outlets and below-median outlets can be attributed to journalist effects.

The results in Figure 10 are from a restrictive set of authors and estimates first moment differences across sets of outlets. Table 3 provides estimates from the full sample and utilizes the full variance decomposition by computing $\hat{\text{cov}}(\hat{\theta}^d, \hat{\theta}^o)$.

Our baseline estimate (estimator $\text{cov}(\hat{\theta}^d, \hat{\theta}^o)$ in Column 1 of Table 3) suggests 16 percent of the outlet-level variation in slant can be explained by journalist effects. We can compare this to the estimated 20 percent of observed variation in outlet-level average slant that is explained by consumer demand in Gentzkow and Shapiro (2010). Taken at face value, our results suggest journalist effects are nearly as important as consumer demand in explaining the variation in observed slant at the outlet level.

7 Conclusion

This paper makes three primary contributions.

First, we provide an important advance in measuring media slant—with our fine-tuned RoBERTa model achieving a significant improvement in measuring human-perceived political slant relative to prior methods—and document new stylized facts regarding media slant.

Second, we provide novel evidence on the role journalists play in determining the political slant of the news they produce. Our estimates reject the hypothesis that journalists have zero ideological preferences over the content they produce and imply that 16 percent of observed variation in slant across outlets can be explained by journalist preferences.

Third, to the extent our results speak to worker preferences over the slant they produce, they provide evidence on the importance of non-pecuniary worker preferences in driving sorting across firms.

There are several caveats to our analysis. First, our analysis requires strong assumptions about the selection of journalists across outlets. Violations of these assumptions will bias our estimates. Second, our notion of the role of journalists focuses on journalists as individuals. Group dynamics or the spillover effect of *other* journalists, such as those discussed by Bari Weiss (see Introduction), would be attributed to the outlet. If these dynamics are important, our estimates would be understating the role of journalists. Lastly, measurement error in our measure of slant may also affect our estimates. Future research should continue to examine the effects of these potential limitations on our results.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. High Wage Workers and High Wage Firms. *Econometrica.* 67(2): 251–333.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Perez. 2021. Automated Linking of Historical Data. *Journal of Economic Literature*. 59(3): 865-918.

Baum, Matthew A. and Tim Groeling. 2008. New Media and the Polarization of American Political Discourse. *Political Communication*. 25(4): 345–365.

Barbera, Pablo and Gaurav Sood. 2018. Follow Your Ideology: Measuring Media Ideology on Social Networks. *Working Paper*.

Baron, David P. 2006. Persistent Media Bias. *Journal of Public Economics*. 90(1-2): 1–36.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. A Distributional Framework for Matched Employer Employee Data. *Econometrica*. 87(3): 699–739.

Bonica, Adam. 2013. Ideology and Interests in the Political Marketplace. *American Journal of Political Science*. 57(2): 294–311.

Boxell, Levi. 2021. Slanted images: Measuring nonverbal media bias during the 2016 election. *Political Science Research and Methods*. Conditionally Accepted.

Brenan, Megan. 2020. Americans Remain Distrustful of Mass Media. *Gallup*. Accessed at https://news.gallup.com/poll/321116/americans-remain-distrustful-mass-media.aspx on April 15, 2022.

Budak, Ceren, Sharad Goel, and Justin M. Rao. 2016. Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly.* 80(S1): 250–271.

Cagé, Julia, Mortiz Hengel, Nicolas Hervé, and Camille Urvoy. 2022. Hosting Media Bias: Evidence from the Universe of French Broadcasts, 2002-2020. *Working Paper.*

Colonnelli, Emanuele, Valdemar Pinho Neto, and Edoardo Teso. 2021. Politics at Work. *Working Paper*.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. https://arxiv.org/abs/1810.04805.

Faris, Robert M., Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler. 2017. Partisanship, propaganda, and disinformation: Online media and the 2016 U.S. presidential election. *Berkman Klein Center for Internet & Society Research Paper.*

Gentzkow, Matthew and Jesse M. Shapiro. 2010. What Drives Media Bias? Evidence from U.S. Daily Newspapers. *Econometrica*. 78(1): 35–71.

Gentzkow, Matthew, Jesse M. Shapiro, and Matt Taddy. 2019. Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. *Economet-*

rica. 87(4): 1307–1340.

Gift, Karen and Thomas Gift. 2015. Does Politics Influence Hiring? Evidence from a Randomized Experiment. 37(3): 653–675.

Graves, Lucas, Brendan Nyhan, and Jason Reifler. 2016. Understanding Innovations in Journalistic Practice: A Field Experiment Examining Motivations for Fact-Checking. *Journal of Communication.* 66(1): 102–138.

Groseclose, Tim and Jeffrey Milyo. 2005. A Measure of Media Bias. *Quarterly Journal of Economics*. 120(4): 1191–1237.

Groseclose, Tim. 2011. *Left Turn: How Liberal Media Bias Distorts the American Mind.* St. Martin's Griffin: New York, NY.

Hassell, Hans J.G., John B. Holbein, and Matthew R. Miles. 2020. There is No Liberal Media Bias in which News Stories Political Journalists Choose to Cover. *Science Advances*. 6(14).

Hedblom, Daniel, Brent R. Hickman, and John A. List. 2019. Toward an Understanding of Corporate Social Responsibility: Theory and Field Experimental Evidence. *Working Paper*.

Joulin, Armand, Edouardo Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of tricks for efficient text classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers.*

Kline, Patrick, Raffeale Saggio, Mikkel Solvsten. 2020. Leave-out Estimation of Variance Components. *Econometrica*. 88(5): 1859–1898.

Liu, Yinhan and others. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. https://arxiv.org/abs/1907.11692v1.

Martin, Gregory and Ali Yurukoglu. 2017. Bias in Cable News: Persuasion and Polarization. *American Economic Review*. 107(9): 2565–2599.

Martin, Gregoy and Joshua McCrain. 2019. Local News and National Politics. *American Political Science Review*. 113(2): 372–384.

Povich, Elaine. 1996. Partners and Adversaries: The Contentious Connection Between Congress and the Media. *Freedom Forum*. Referenced in https://www.mrc.org/exhibit-1-9-washington-bureau-chiefs-and-correspondents.

Sauter, Van Gordon. 2020. The 'Liberal Leaning' Media Has Passed Its Tipping Point. *Wall Street Journal*. Accessed at https://www.wsj.com/articles/the-liberal-leaning-media-has-passed-its-tipping-point-11590430876 on April 15, 2022.

Shapiro, Jesse M. 2016. Special Interests and the Media: Theory and an Application to Climate Change. *Journal of Public Economics*. 114: 91–108.

Smith, Noah. 2021. Interview: Matt Yglesias, author and blogger. *Noahpinion*. Accessed at https://noahpinion.substack.com/p/interview-matt-yglesias-author-and on April 15, 2022.

Sorkin, Isaac. 2018. Ranking Firms Using Revealed Preferences. *Quarterly Journal of Economics*.

133(3): 1331–1393.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. https://arxiv.org/abs/1706.03762.

Weaver, David H., Lars Willnat, G. Cleveland Wilhoit. 2019. The American Journalist in the Digital Age: Another Look at U.S. News People. *Journalism & Mass Communication Quarterly*. 96(1): 101–130.

Weiss, Bari. 2020. Resignation Letter. Accessed at https://www.bariweiss.com/resignation-letter on April 15, 2022.

Wemple, Erik. 2017. Dear Mainstream Media: Why so Liberal? *Washington Post*. Accessed at https://www.washingtonpost.com/blogs/erik-wemple/wp/2017/01/27/dear-mainstream-media-why-so-liberal/ on April 15, 2022.

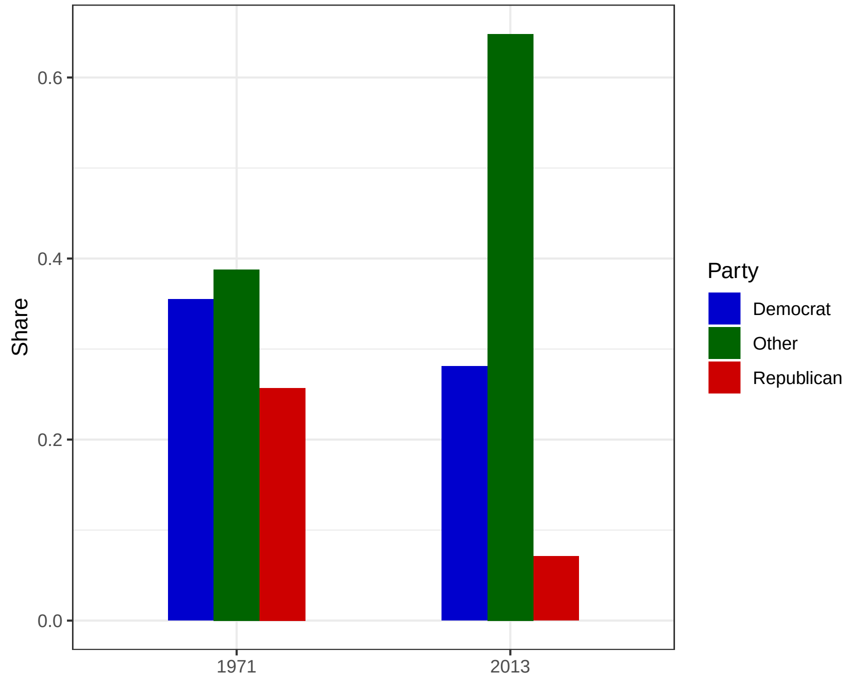
Wu, Yonghui and others. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. https://arxiv.org/abs/1609.08144.

Xu, Guosong. 2022. Friends in Media. *Working Paper*.

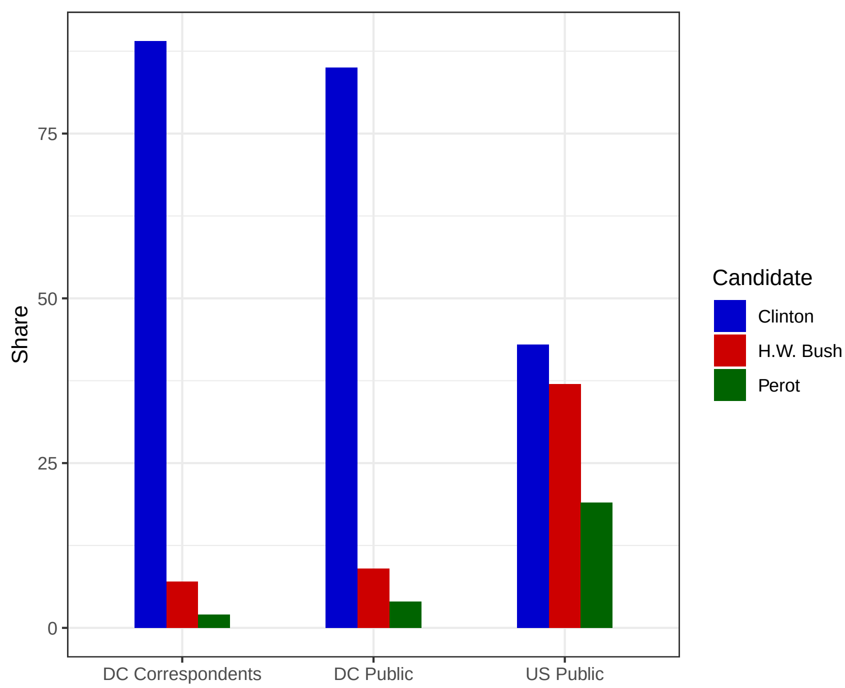
Zhu, Ykun, Ryan Kiros, Richard Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Anotonio Torralba, and Sanja Fidler. 2015. Aligning books and moves: Towards story-like visual explanations by watching movies and reading books. https://arxiv.org/abs/1506.06724.

Figure 1: Partisan Distribution of Journalists

Panel A: Surveyed Journalist's Self-Reported Partisan Identification (Weaver et al. 2019)



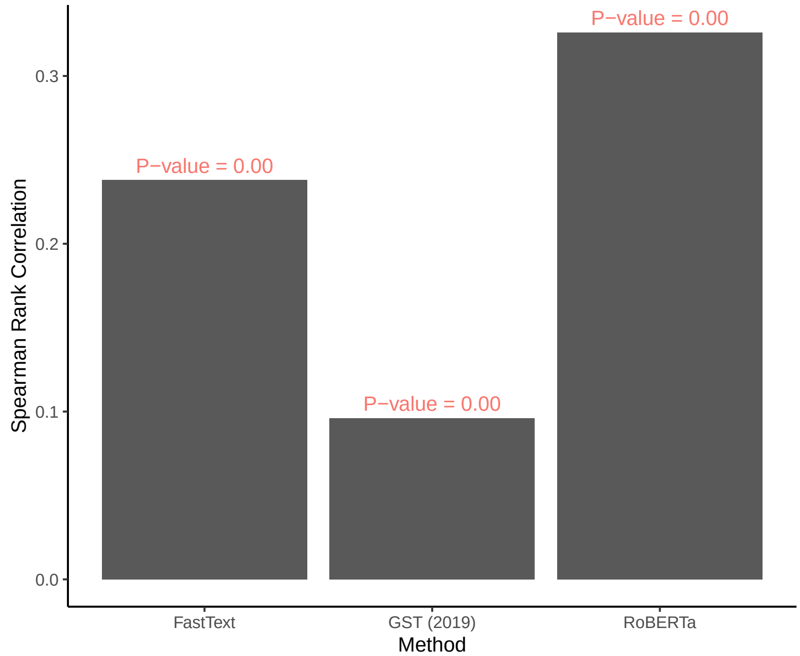
Panel B: Surveyed Washington Correspondent's Self-Reported Vote Choice in 1992 (Povich 1996)



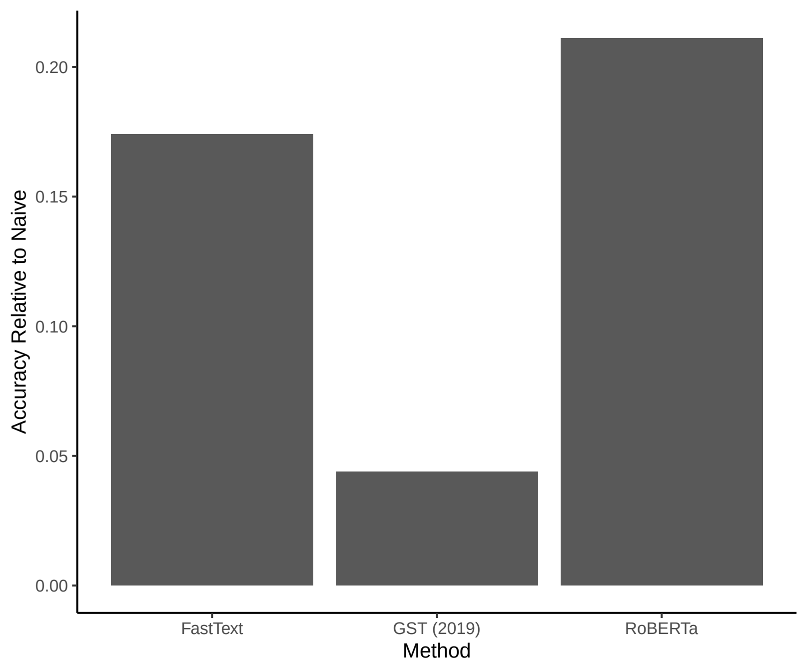
Notes: Panel A reports the share of surveyed journalists in 1971 and 2013 that self-reported as a Democrat (blue), a Republican (red), or as non-partisan or a third party (green). Data is taken from Weaver et al. (2019). Panel B reports the self-reported vote choices of Washington Correspondents from Povich (1996), as well as information on vote choices for the District of Columbia and the US population as a whole.

Figure 2: Comparing to Human Perceived Slant

Panel A: Non-parametric Rank Correlation

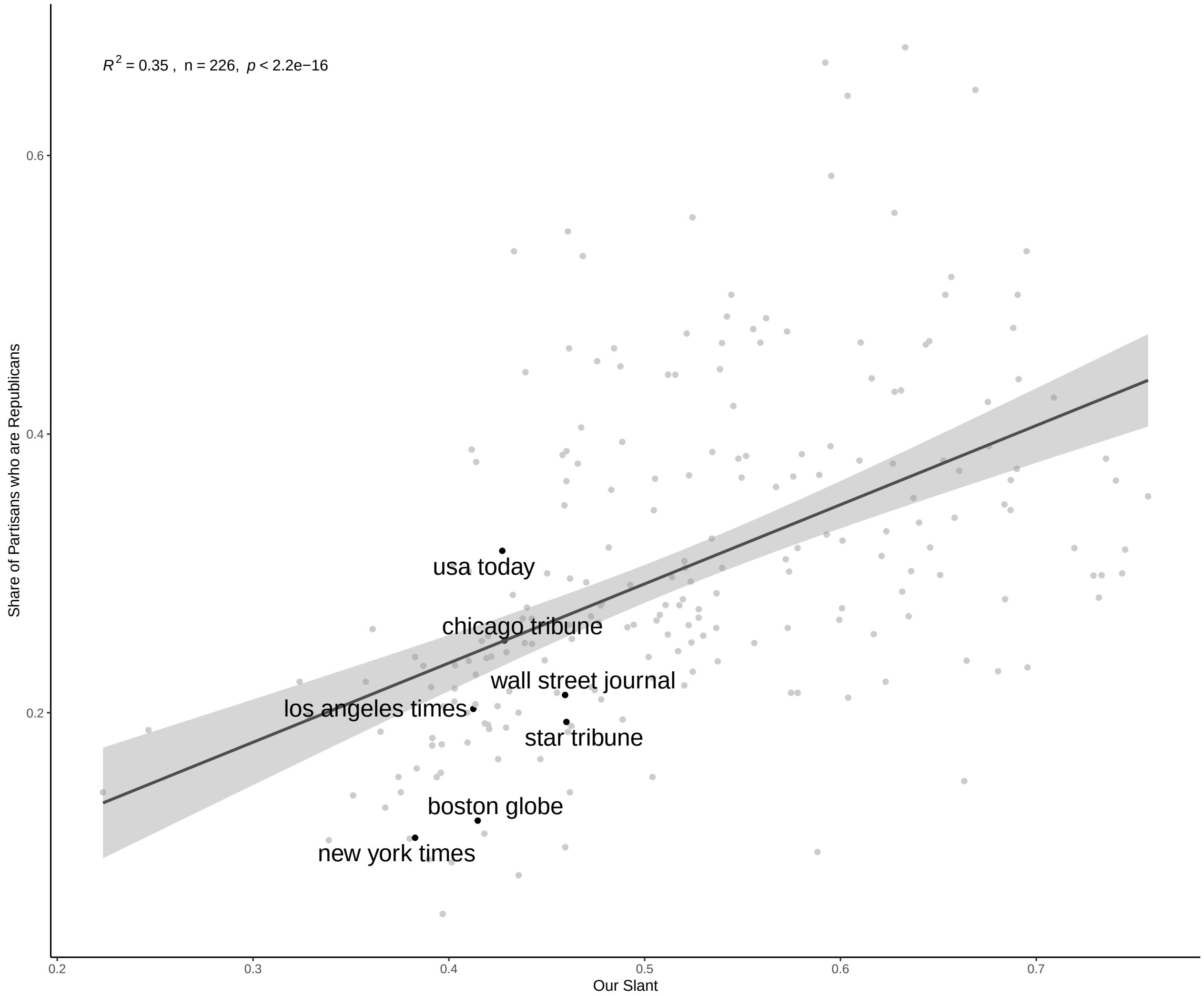


Panel B: Accuracy among Clearly Partisan Articles



Notes: Panel A plots the rank correlation of the Budak et al. (2016) perceived slant measure for each article as defined in Appendix Table B1 versus the model-based slant measures using all articles. Panel B plots the accuracy of each model-based method in partitioning the dataset in half between more liberal and more conservative articles relative to a naive algorithm. In Panel B, the sample is restricted to articles for which both MTurk coders gave the same relative rating and for which this relative rating was non-zero.

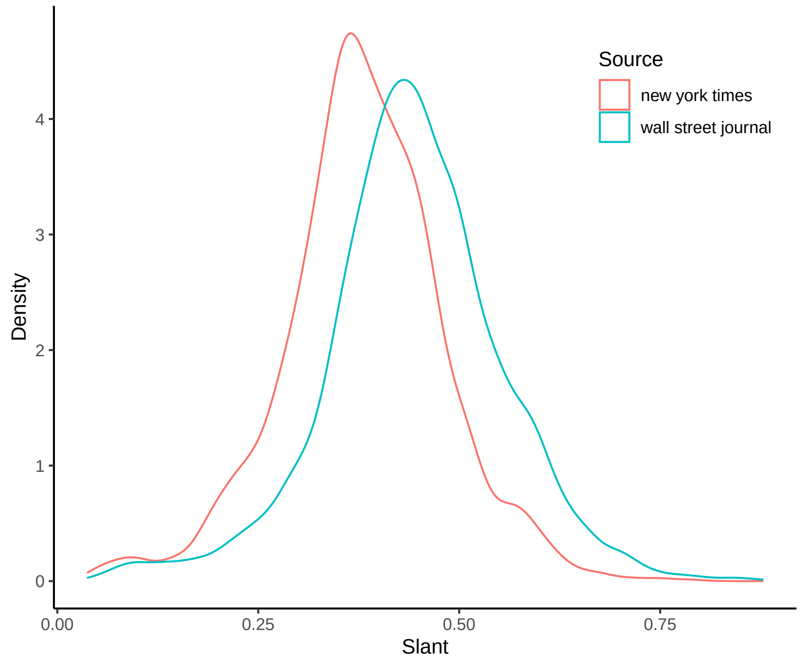
Figure 3: Average Outlet Slant and Journalist Partisanship



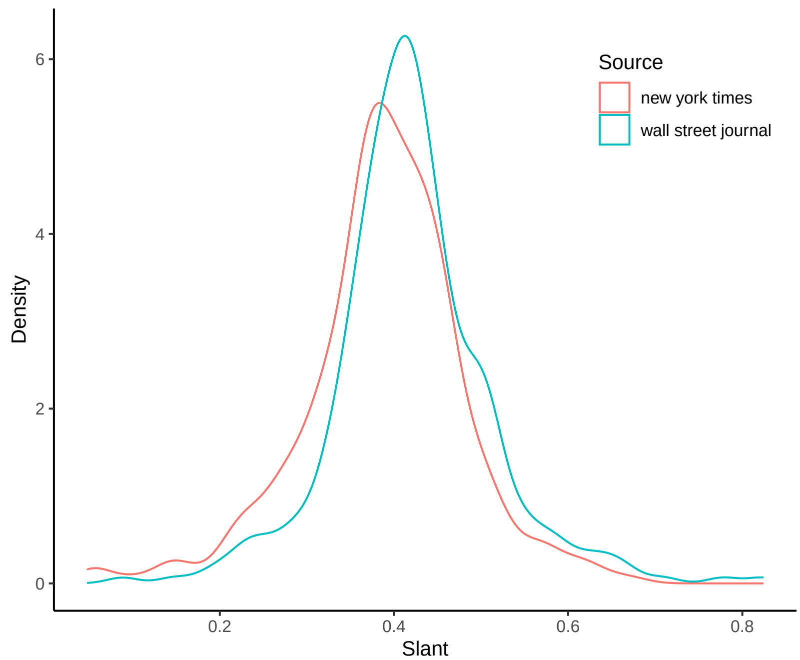
Notes: Figure plots the relationship between the average estimated slant at an outlet (x-axis) with the share of matched partisan journalists who are Republicans. Slant ranges from 0 to 1, with larger values indicating content estimated to be more conservative.

Figure 4: Authors at *The New York Times* versus *The Wall Street Journal*

Panel A: All Authors

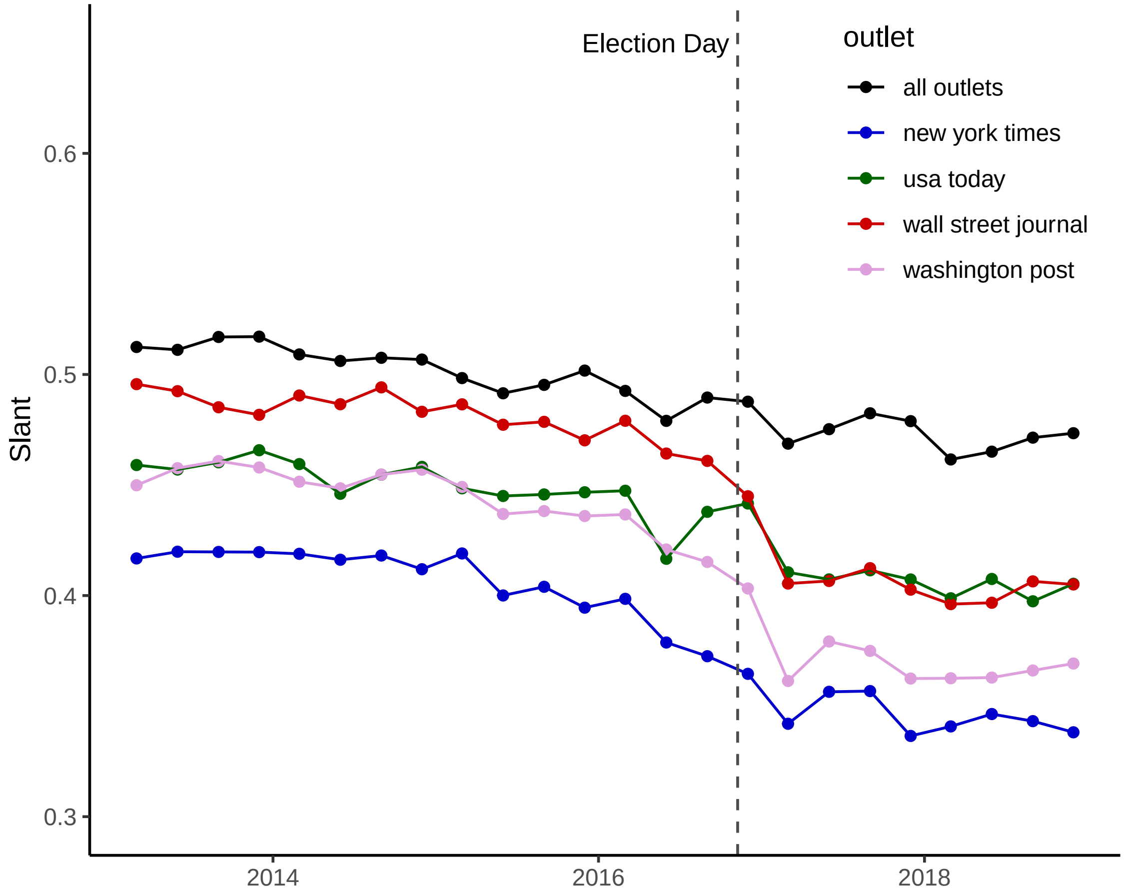


Panel B: Movers Between NYT and WSJ



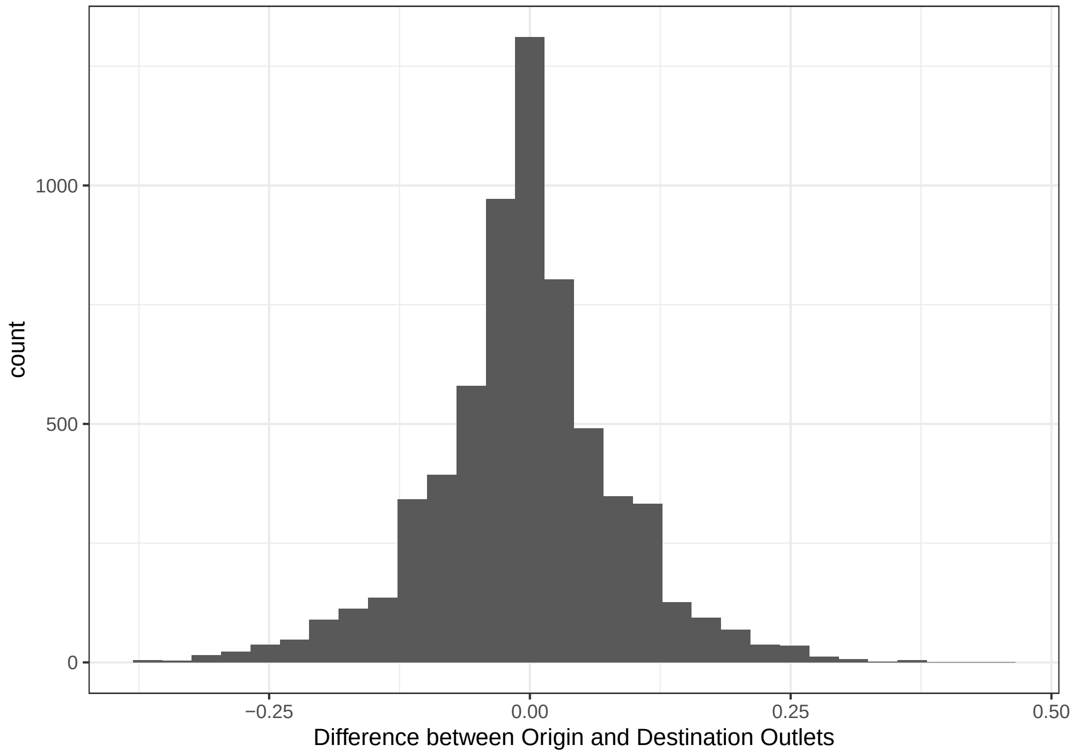
Notes: Figure plots the distribution, across authors, of the average slant of each author's article as measured by our RoBERTa model of slant for each outlet. Panel A includes all authors who appear at a given outlet. Panel B restricts attention to authors who have written at both *The New York Times* and *The Wall Street Journal*. Slant ranges from 0 to 1, with larger values indicating content estimated to be more conservative.

Figure 5: Trends in Slant



Notes: Figure plots the trends in estimated slant over time, aggregated by quarter, for all outlets (black) as well as selected outlets (various colors) in our main analysis sample. The dashed vertical line is Election Day 2016. Slant ranges from 0 to 1, with larger values indicating content estimated to be more conservative.

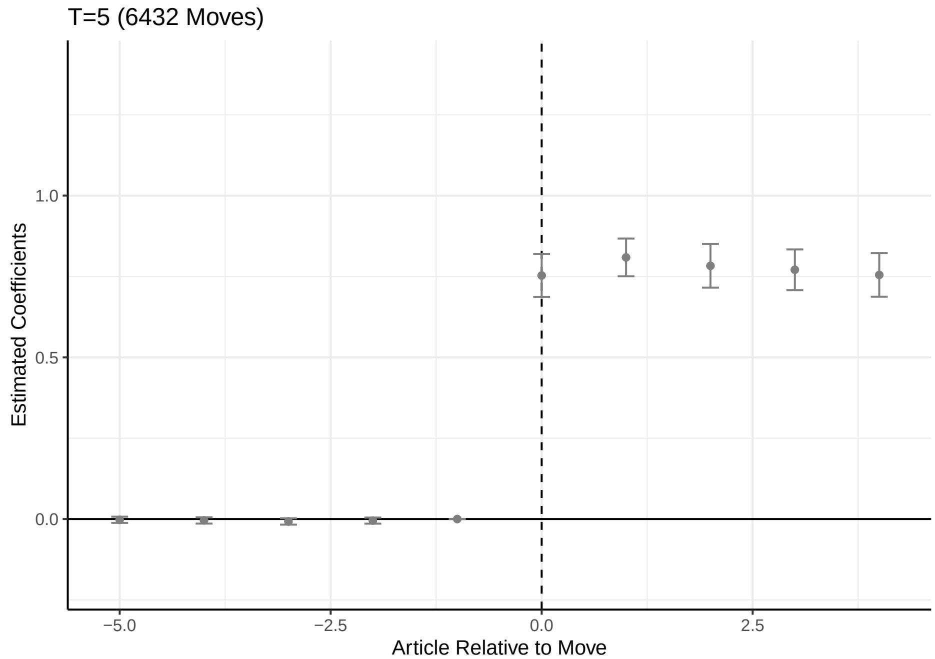
Figure 6: Distribution of Origin-Destination Outlet Slant Scores in Event Study Sample ($T = 5$)



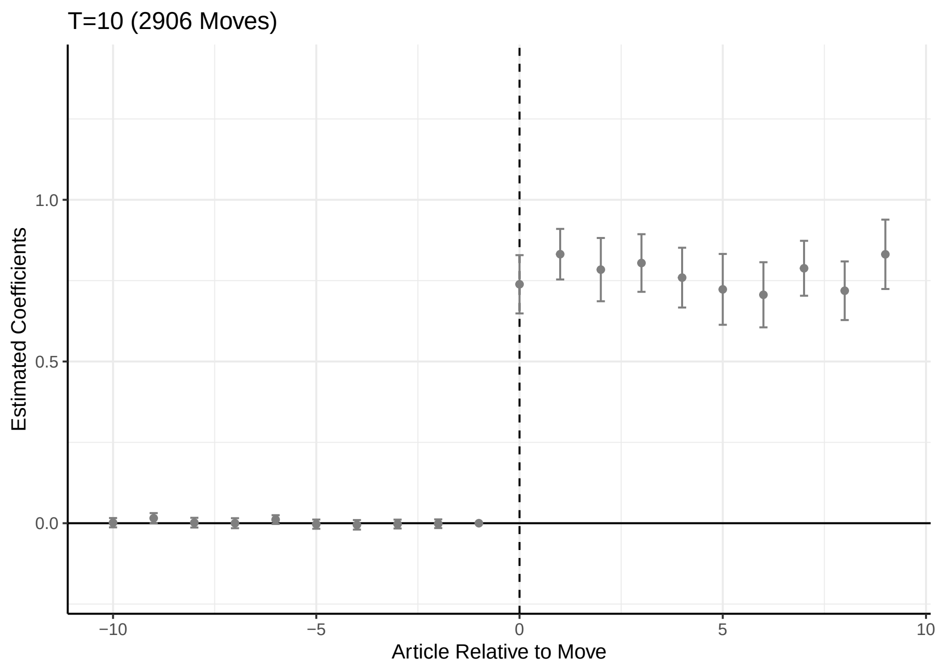
Notes: Figure plots the distribution of moves $\bar{x}_{d(i)} - \bar{x}_{o(i)}$ for consistent movers in our $T = 5$ estimation sample.

Figure 7: Event Study Changes in Slant Around Author Moves

Panel A: $T = 5$



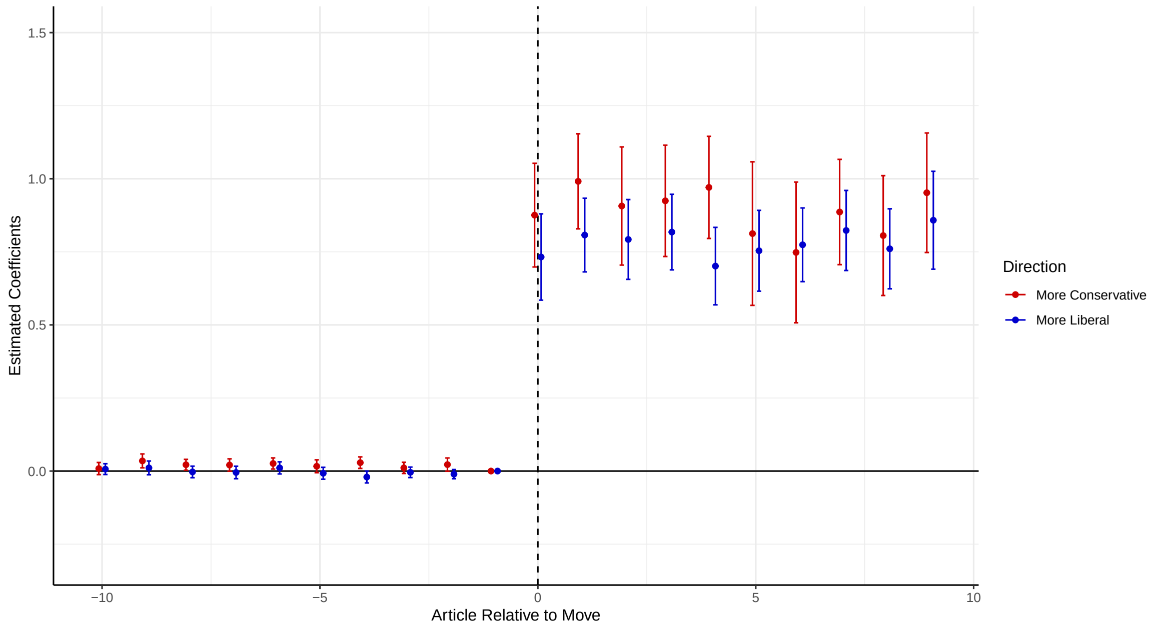
Panel B: $T = 10$



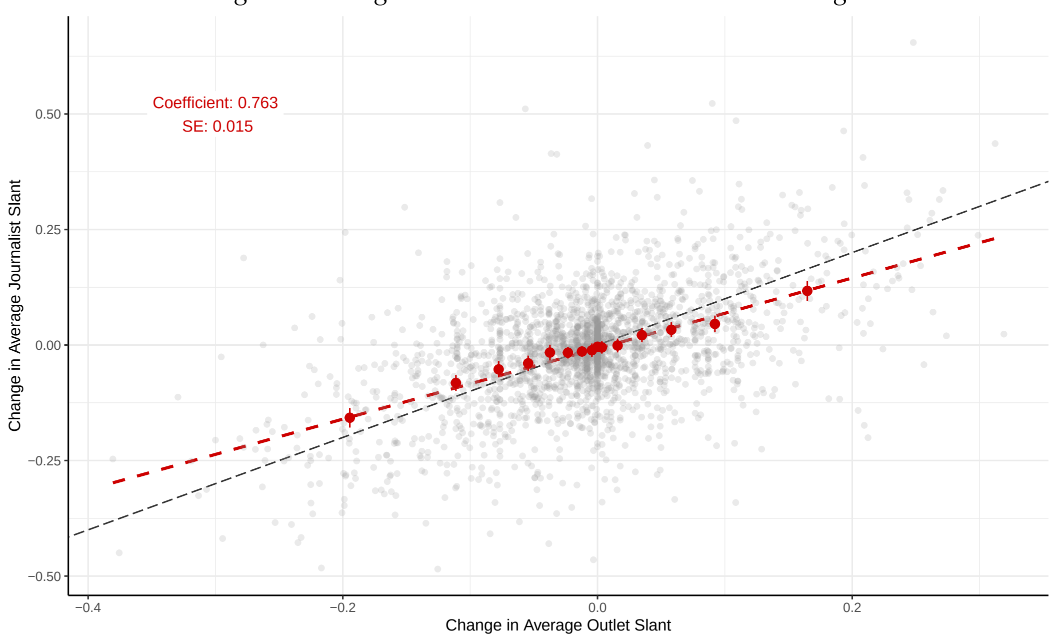
Notes: Figure plots the estimated event study coefficients $\hat{\rho}_l$ for the $T = 5$ estimation sample (Panel A) and the $T = 10$ estimation sample (Panel B). Event study coefficients estimate the moving author's change in slant relative to their slant at $t = -1$.

Figure 8: Exploring Potential Asymmetry

Panel A: Event Study by Direction of Move

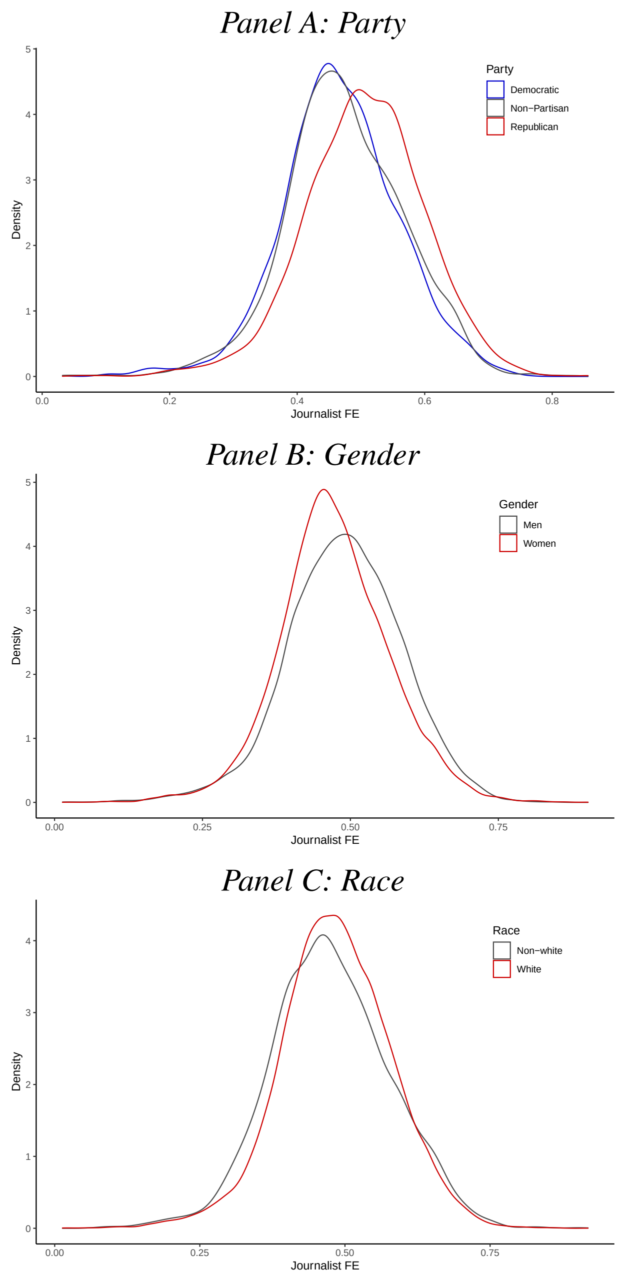


Panel B: Predicting the Change in Journalist Slant with the Change in Outlet Slant



Notes: Panel A plots the event study coefficients for the $T = 10$ estimation sample when the model is estimated separately among the set of movers who move to a more conservative outlet (e.g., $\bar{x}_{d(i)} - \bar{x}_{o(i)} > 0$) in red and the set of movers who move to a more liberal outlet (e.g., $\bar{x}_{d(i)} - \bar{x}_{o(i)} < 0$) in blue. Panel B plots the (bin)scatter plot for the $T = 10$ estimation sample with $\bar{x}_{d(i)} - \bar{x}_{o(i)}$ on the x-axis and $\frac{1}{T+1}\sum_{l\geq 0}\hat{x}_{il} - \frac{1}{T}\sum_{l<0}\hat{x}_{il}$ on the y-axis. Observations are weighted by $(\bar{x}_{d(i)} - \bar{x}_{o(i)})^2$.

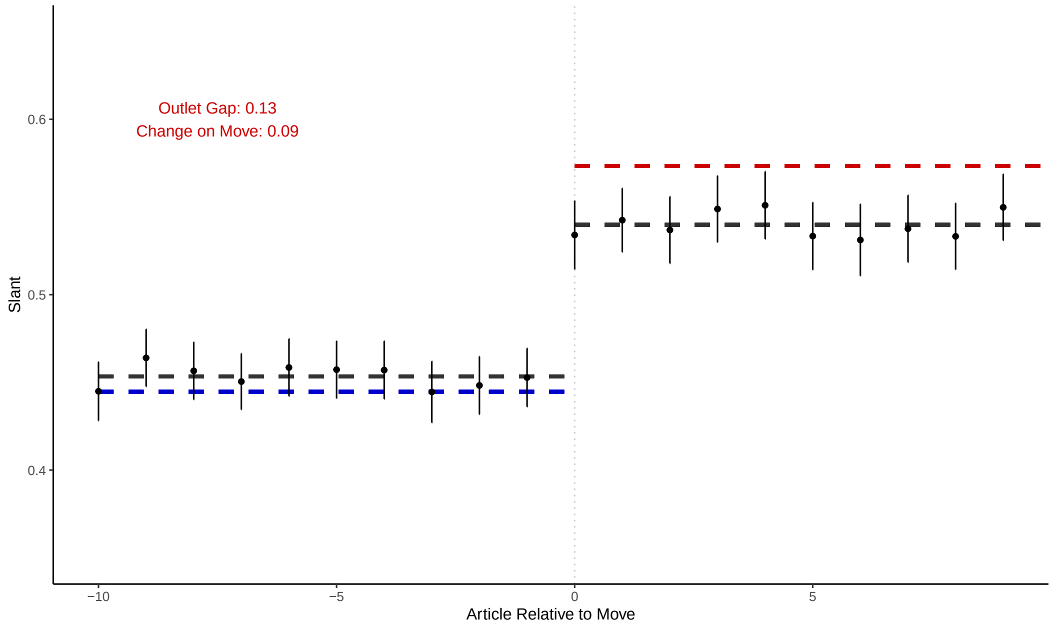
Figure 9: Heterogeneity in Journalist AKM Fixed Effects



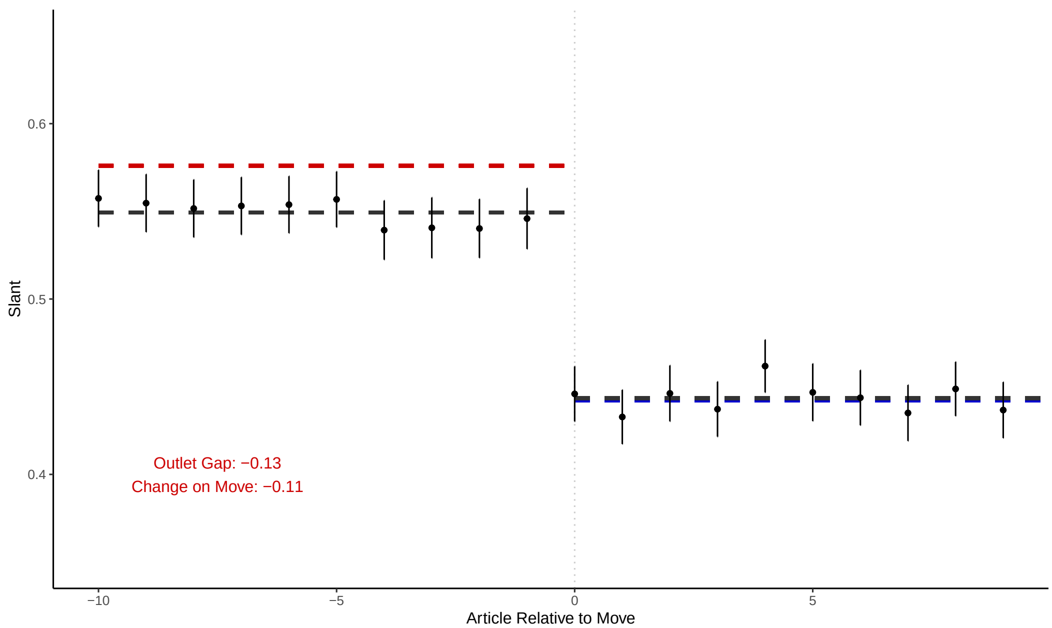
Notes: Each panel plots the distribution in estimated AKM journalist fixed effects across various sub-groups of journalists. Panel A examines heterogeneity across journalists by their partisan affiliation. Panel B examines heterogeneity by gender, as predicted by the author's name. Panel C examines heterogeneity by race, as predicted by the author's name.

Figure 10: Decomposition Event Study

Panel A: Below-Median to Above-Median Movers



Panel B: Above-Median to Below-Median Movers



Notes: Panel A plots the average values of $\bar{x}_{j'}$, $\bar{x}_j$, $x_{ij'}$, and x_{ij} for the set of movers who move from below-median slant outlets to above-median slant outlets. The dashed blue (red) line is the average $\bar{x}_j$ for below-median (above-median) slant outlets. The grey points and associated confidence intervals are the average article slant for the tth article relative to the move for each journalist in the corresponding set. The dashed grey lines are the average across the grey points before and after the move respectively. Panel B is the same as Panel A, except using the set of movers from above-median to below-median outlets.

Table 1: Summary Statistics

	Variable					
	Articles	Authors	Slant	Slant (Stayers)	# of Movers	# of Stayers
Boston Globe	184340	1795	0.41	0.40	1371	424
Chicago Tribune	147476	4912	0.43	0.48	4703	209
Los Angeles Times	117037	2117	0.41	0.36	2080	37
New York Times	175459	1969	0.38	0.40	1665	304
USA Today	104441	2519	0.43	0.53	2503	16
Wall Street Journal	445307	2773	0.46	0.47	1564	1209
Washington Post	135171	2069	0.41	0.43	1975	94
All Outlets	7873626	31985	0.49	0.50	20597	11388

Notes: Table shows various summary statistics for our cleaned sample of ProQuest data.

Table 2: AKM Decompositions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sample:	All	Dropping Stayers (Full)	Dropping Stayers (Large)	Max 2 Outlets	Drop 1 Mo. Moves	Top 20 Outlets	$T=5$ Full	$T=5$ Large	$T=10$ Full	Parent Company
Level of Aggregation:	Month	Month	Month	Month	Month	Month	Article	Article	Article	Month
Panel A: Plug-in Estimator										
	Share of Variance, x_{ijt}									
Journalist, var(γ_i)	0.353	0.313	0.296	0.365	0.354	0.425	0.258	0.264	0.228	0.527
Outlet, var(α_j)	0.244	0.256	0.323	0.254	0.249	0.128	0.160	0.263	0.185	0.056
Sorting, $2\times\text{cov}(\gamma_i,\alpha_j)$	0.081	0.088	0.044	0.060	0.075	0.006	0.065	-0.030	0.042	0.063
	Correlation									
Sorting, cor(γ_i,α_j)	0.137	0.156	0.072	0.099	0.127	0.013	0.159	-0.056	0.103	0.183
Panel B: Leave-Out Estimator (Kline et al. 2020)										
	Share of Variance, x_{ijt}									
Journalist, var(γ_i)	0.299	0.198	0.148	0.297	0.299	0.384	0.157	0.136	0.139	0.484
Outlet, var(α_j)	0.226	0.239	0.290	0.219	0.227	0.112	0.143	0.212	0.163	0.051
Sorting, $2\times\text{cov}(\gamma_i,\alpha_j)$	0.116	0.113	0.092	0.128	0.119	0.038	0.085	0.029	0.070	0.072
	Correlation									
Sorting, cor(γ_i,α_j)	0.222	0.259	0.221	0.251	0.228	0.091	0.282	0.086	0.232	0.229

Notes: Table reports AKM variance decomposition exercises using the plug-in estimator (Panel A) and the Kline et al. (2020) leave-out estimator (Panel B) with various samples. Both panels are restricted to the leave-out connected set of outlets. Data is aggregated to the author-source-month level unless otherwise noted. Column (1) uses the full sample. Column (2) drops journalists that never switch outlets. Column (3) is the same as Column (2) except that it also restricts to journalists whose range in outlet slant is greater than 0.1. Column (4) drops journalists that appear in more than two outlets. Column (5) restricts to tenures of longer than one month in duration. Column (6) restricts to the top 20 outlets according to weekend circulation. Column (7) uses the $T=5$ event study sample at the article level. Column (8) is the same as Column (7) except that it restricts to moves where $|\bar{x}_{d(i)} - \bar{x}_{o(i)}| > .1$. Column (9) uses the $T=10$ event study sample at the article level. Column (10) treats all outlets owned by a single "Parent Company" as the same outlet and restricts attention to outlets for which we observe the parent company.

Table 3: Outlet-level Journalist Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sample:	All	12+ Months	Max 2 Outlets	Drop 1 Mo. Moves	Top 20 Outlets	$T=5$ Sample	$T=10$ Sample	Parent Company
Estimator	Share of Variance, $\bar{x}_j$							
$\mathrm{cov}(\hat{\theta}^d, \hat{\theta}^o)$	0.159	0.169	0.067	0.150	0.311	0.331	0.086	0.425
$\mathrm{cov}(\hat{\theta}^d, \hat{\theta}^o)$, unweighted	0.239	0.220	0.072	0.204	0.240	0.201	0.052	0.560
$\mathrm{var}(\hat{\theta}^o)$	0.180	0.191	0.560	0.183	0.384	0.633	0.678	0.497
$\mathrm{var}(\hat{\theta}^d)$	0.180	0.184	0.557	0.183	0.289	0.869	0.895	0.407
	Correlation							
$\mathrm{cor}(\hat{\theta}^d, \hat{\theta}^o)$	0.885	0.903	0.120	0.820	0.933	0.447	0.111	0.945
$\mathrm{cor}(\hat{\theta}^d, \hat{\theta}^o)$, unweighted	0.870	0.866	0.101	0.809	0.947	0.447	0.111	0.953
	Test Equality							
Coef.	0.870	0.898	0.122	0.797	0.847	0.387	0.145	0.992
	(0.033)	(0.028)	(0.096)	(0.041)	(0.108)	(0.006)	(0.007)	(0.043)
P-value (Test Coef = 1)	0.000	0.000	0.000	0.000	0.174	0.000	0.000	0.862
	Sample Size							
Observations	19516	17422	1922	10863	3104	4342	1721	11542
Journalists	5513	4235	1922	5075	1496	2892	1432	3503
Outlets	284	279	228	278	18	259	217	35

Notes: Table reports our estimates of the contribution of outlet-level journalist effects to the observed variation in average outlet slant using various estimators. All estimates are weighted by the number of unique authors at the outlet unless otherwise noted. Column (1) uses the full sample. Column (2) restricts to journalists with observations in at least 12 months. Column (3) drops journalists that appear in more than two outlets. Column (4) restricts to tenures of longer than one month in duration. Column (5) restricts to the top 20 outlets according to weekend circulation. Column (6) uses the $T = 5$ event study sample at the article level. Column (7) uses the $T = 10$ event study sample at the article level. Column (8) treats all outlets owned by a single "Parent Company" as the same outlet and restricts attention to outlets for which we observe the parent company.

A Theoretical Appendix

A.1 Proof of Proposition 1

The firm's first-order condition with respect to ω is

$$M_i(\omega,\cdot) = \Delta U_{j'}(x,\omega)\frac{\partial}{\partial \omega}M_i(\omega,\cdot)$$

which we can write as

$$v_{j'} - \beta_{j'}(x-\delta_{j'}-\xi_{ij'})^2 - \omega = 1 + \frac{\exp\left(\omega - \beta_i(x-\delta_i)^2\right)}{\exp(\omega_o) + \exp\left(\omega_0 - \beta(x_0-\delta_i)^2\right)}$$

and, with respect to x, the first-order condition is

$$-M_i(x,\cdot) \times \frac{\partial}{\partial x}R_{j'}(A_{j'} \cup \{x\}) = \Delta U_{j'}(x,\omega)\frac{\partial}{\partial x}M_i(x,\cdot)$$

We can combine first-order conditions to get

$$\frac{\partial}{\partial x}R_{j'}(A_{j'} \cup \{x\}) = -\frac{\frac{\partial}{\partial x}M_i(x,\cdot)}{\frac{\partial}{\partial \omega}M_i(\omega,\cdot)}.$$

And noting that $\frac{\partial}{\partial x}R_{j'}(A_j \cup \{x\}) = 2\beta_{j'}(x-\delta_{j'}-\xi_{ij'})$ and $-\frac{\frac{\partial}{\partial x}M_i(x,\cdot)}{\frac{\partial}{\partial \omega}M_i(\omega,\cdot)} = -2\beta_i(x-\delta_i)$, we get

$$\beta_{j'}(x-\delta_{j'}-\xi_{ij'}) = -\beta_i(x-\delta_i)$$

which we can re-arrange as

$$x^*_{ij'} = \frac{\beta_i\delta_i + \beta_{j'}(\delta_{j'}+\xi_{ij'})}{\beta_i+\beta_{j'}}.$$

We can then plug x^* into the first order condition for ω to determine h. **QED.**

A.2 Proof of Proposition 2

Note that the optimal slant, from Proposition 1, is

$$x^*_{ij} = \delta_j + \xi_{ij}$$

and the probability a journalist moves conditional on receiving an offer (x,ω) is

$$M_i(x,x_0,\omega,\omega_0) = \frac{\exp(\omega)}{\exp(\omega_o) + \exp(\omega) + \exp(\omega_0)}$$

when $\beta_i = 0$. If $\mathbb{E}(M_i(\cdot)|\xi_{ij}) = c$ for some $c \in \mathbb{R}$, then $\mathbb{E}(\xi_{ij}|i \in A(j)) = 0$. From our assumptions on the search process, $\mathbb{E}(\omega_0|\xi_{ij}) = a$ and $\mathbb{E}(\omega_o|\xi_{ij}) = b$ for some $a, b \in \mathbb{R}$.

Note that, when $\beta_i = 0$, we can rewrite the firm's first-order condition with respect to ω as

$$v_j - \omega = 1 + \frac{\exp(\omega)}{\exp(\omega_o) + \exp(\omega_0)},$$

which is only a function of variables that are independent of ξ_{ij} (i.e., the previous wage ω_0 and the outside option ω_o).[14] Therefore, the optimal wage offer ω is independent of ξ_{ij} and the probability a journalist accepts the offer is independent of ξ_{ij}, i.e., $\mathbb{E}(M_i(\cdot)|\xi_{ij}) = c$ for some $c \in \mathbb{R}$. Thus, $\mathbb{E}(\xi_{ij}|i \in A(j)) = 0$ and $\mathbb{E}(x^*_{ij}|i \in A(j)) = \delta_j$. **QED.**

A.3 Proof of Proposition 3

Note that

$$\mathbb{E}(\hat{\theta}) = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'(\mathbf{T}\psi - \mathbb{E}(\mathbf{e})).$$

And $\mathbb{E}(e_i|\mathbf{T}) = 0$ from the equal selection assumption. And so,

$$\mathbb{E}(\hat{\theta}) = \frac{\beta}{\beta+1}[\bar{\delta}_{A(j)}, \bar{\delta}_{A(j)}]' = [\psi_j, \psi_j]'.$$

Let $\mathbf{A}$ be defined such that $cov(\hat{\theta}^o, \hat{\theta}^d) = \frac{1}{J-1}\hat{\theta}'\mathbf{A}\hat{\theta}$.[15] The quadratic form implies

$$\mathbb{E}\left(\hat{\theta}'\mathbf{A}\hat{\theta}\right) = \text{trace}(\mathbf{A}\mathbb{V}(\hat{\theta})) + \mathbb{E}(\hat{\theta})\mathbf{A}\mathbb{E}(\hat{\theta}).$$

And so

$$\mathbb{E}\left(\hat{\theta}'\mathbf{A}\hat{\theta}\right) = \text{trace}(\mathbf{A}\mathbb{V}(\hat{\theta})) + (J-1)\mathbb{V}\left(\psi_j\right).$$

Finally, note that $\lim_{N_m \to \infty} \frac{1}{J-1}\text{trace}(\mathrm{A}\mathbb{V}(\hat{\theta})) = 0$. **QED.**

---

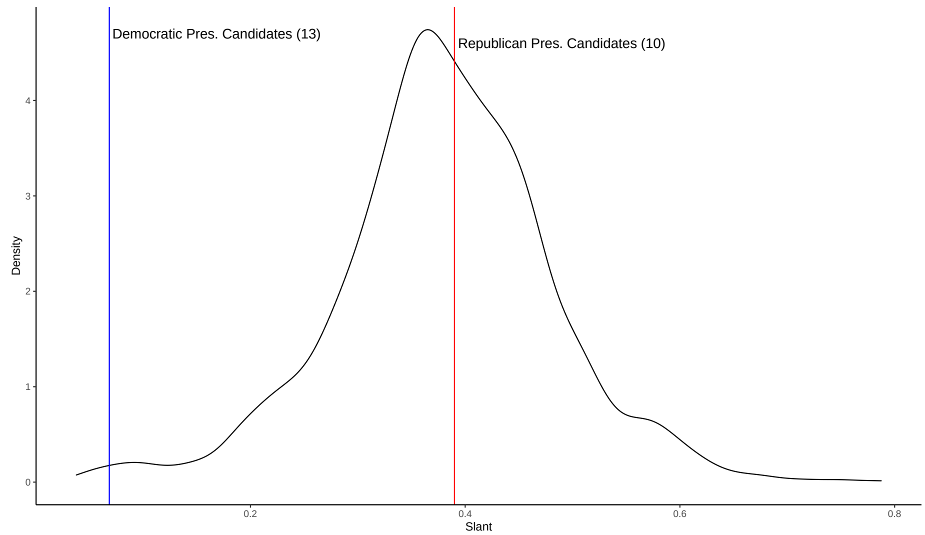
[14] In this case, the optimal wage is $\omega^* = d - W(e^d)$ where $d = \frac{(e^{\omega_0}+e^{\omega_o})(v_j-1)}{1+(e^{\omega_0}+e^{\omega_o})}$ and $W(\cdot)$ is the product log function (see https://reference.wolfram.com/language/ref/ProductLog.html).

[15] Let A_0 be a $J \times J$ matrix of zeros. Let A_1 be a $J \times J$ matrix with $1 - \frac{1}{J}$ on the diagonal and $-\frac{1}{J}$ on the off-diagonals. Then $A = \begin{bmatrix} A_0 & A_1 \\ A_0 & A_0 \end{bmatrix}$.

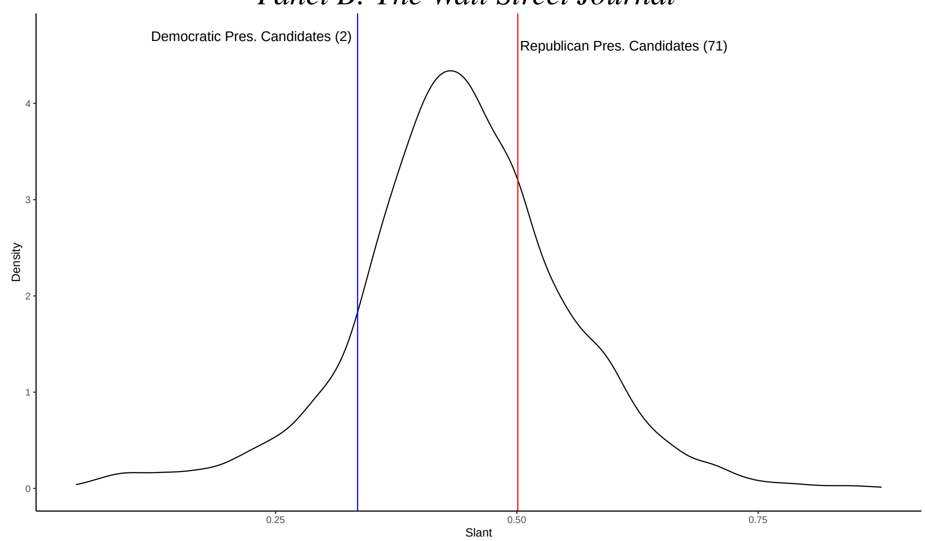
B Empirical Appendix

Appendix Figure B1: Politician Op-Eds

Panel A: The New York Times



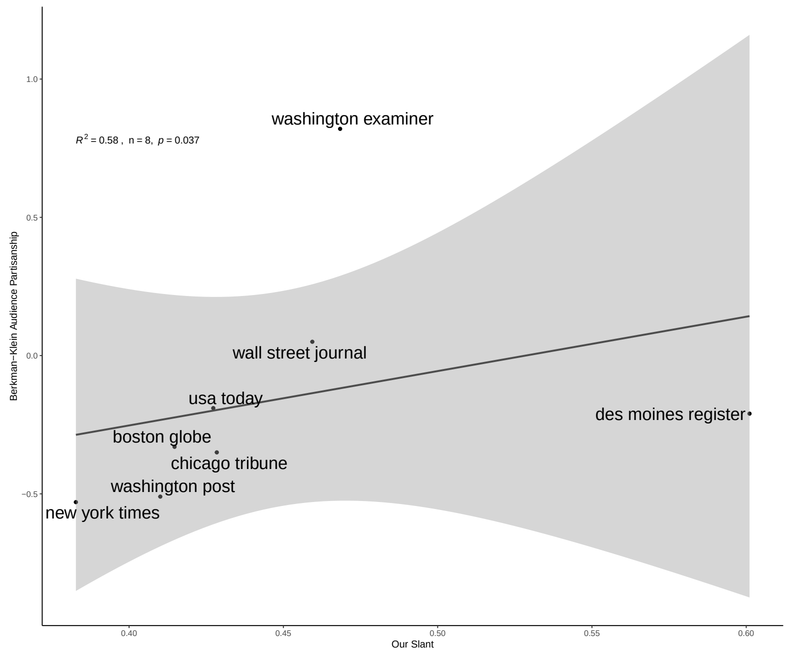
Panel B: The Wall Street Journal



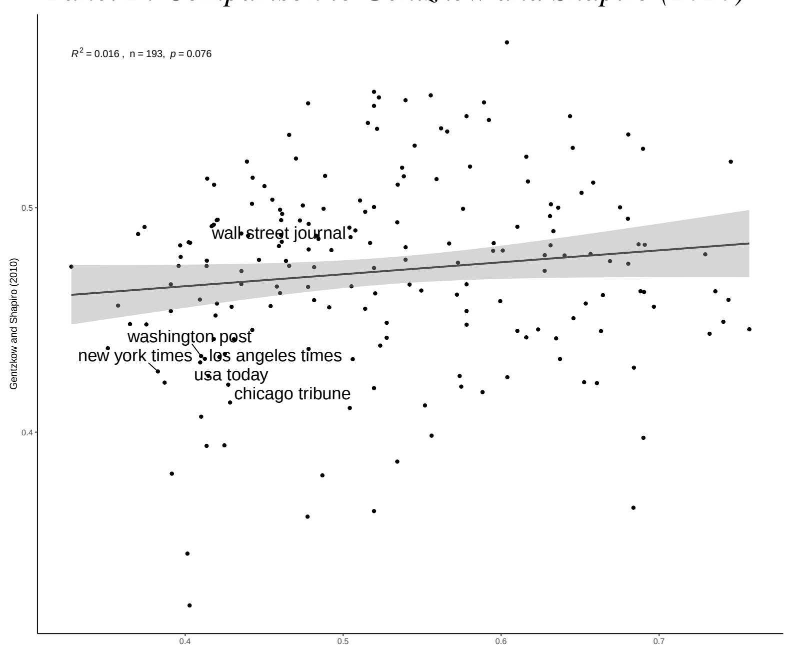
Notes: Panel A plots the distribution, across authors, of the average slant of each author's article as measured by our RoBERTa model of slant for *The New York Times* for our main analysis dataset. The vertical lines display the average slant for major Democratic presidential candidates (blue) and major Republican presidential candidates (red) in 2016 across articles written by those authors in *The New York Times*. Panel B is the same as Panel A, except for *The Wall Street Journal.* The number of articles in each set is indicated in parentheses. The politician articles include articles outside the main analysis dataset.

Appendix Figure B2: Comparison to Other Measures of Slant or Consumer Partisanship

Panel A: Comparison to Berkman-Klein

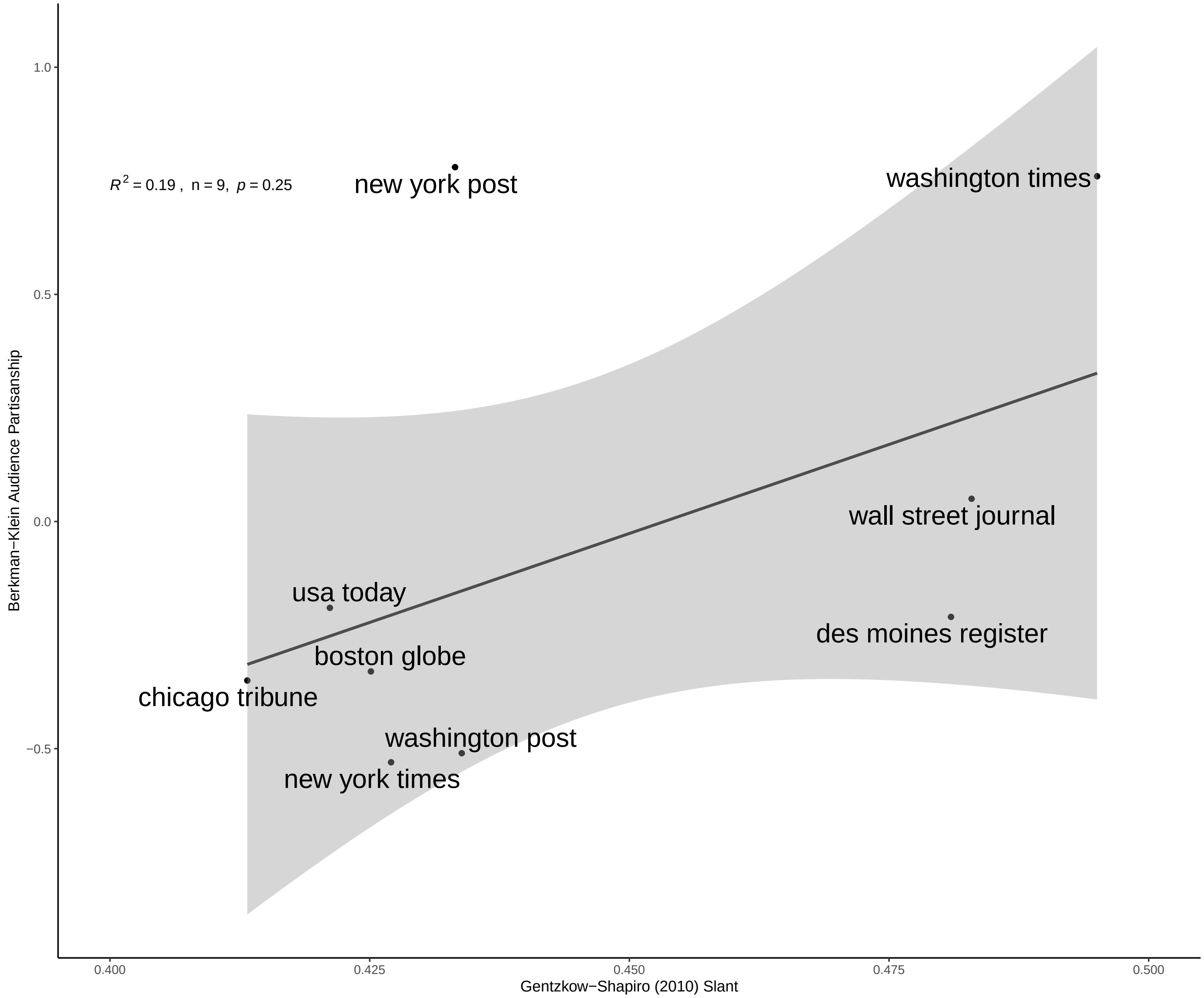


Panel B: Comparison to Gentzkow and Shapiro (2010)



Notes: Panel A plots the comparison of our measure of slant (x-axis) to a measure of audience partisanship from the Berkman-Klein study on the 2016 election (y-axis). Panel B plots the outlet-level slant as measured in Gentzkow and Shapiro (2010) on the x-axis, and the average slant across articles as measured by our RoBERTa model of slant. Each point represents a single outlet. We match outlets between the ProQuest data and the Gentzkow and Shapiro (2010) data using a fuzzy string matching algorithm. We exclude one outlier with a negative slant score in the Gentzkow and Shapiro (2010) data.

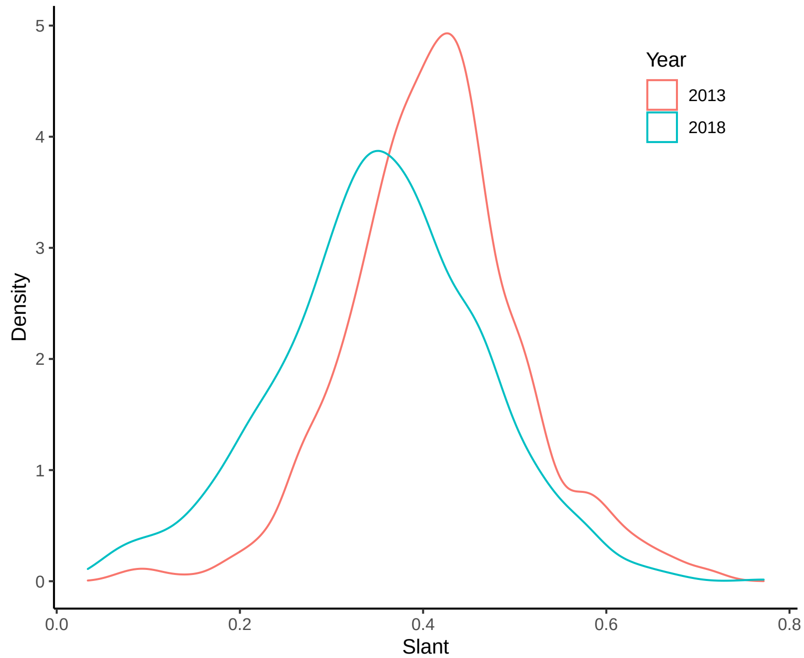
Appendix Figure B3: Comparing Gentzkow and Shapiro (2010) to Berkman-Klein Partisanship



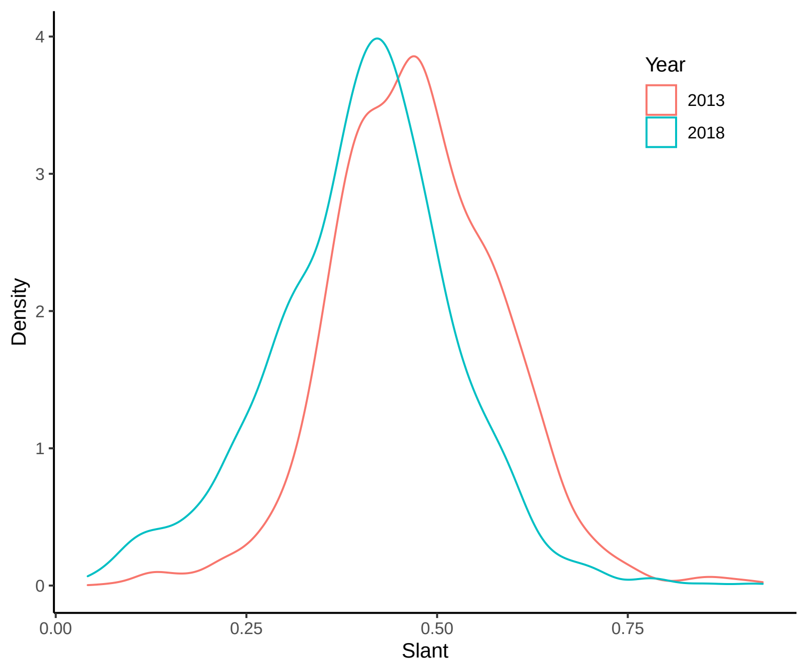
Notes: Figure plots the outlet-level slant as measured in Gentzkow and Shapiro (2010) on the x-axis and a measure of audience partisanship from the Berkman-Klein 2016 election study on the y-axis. Each point represents a single outlet. We exclude one outlier with a negative slant score in the Gentzkow and Shapiro (2010) data.

Appendix Figure B4: Authors at *The New York Times* and *The Wall Street Journal* — 2013 vs. 2018

Panel A: New York Times



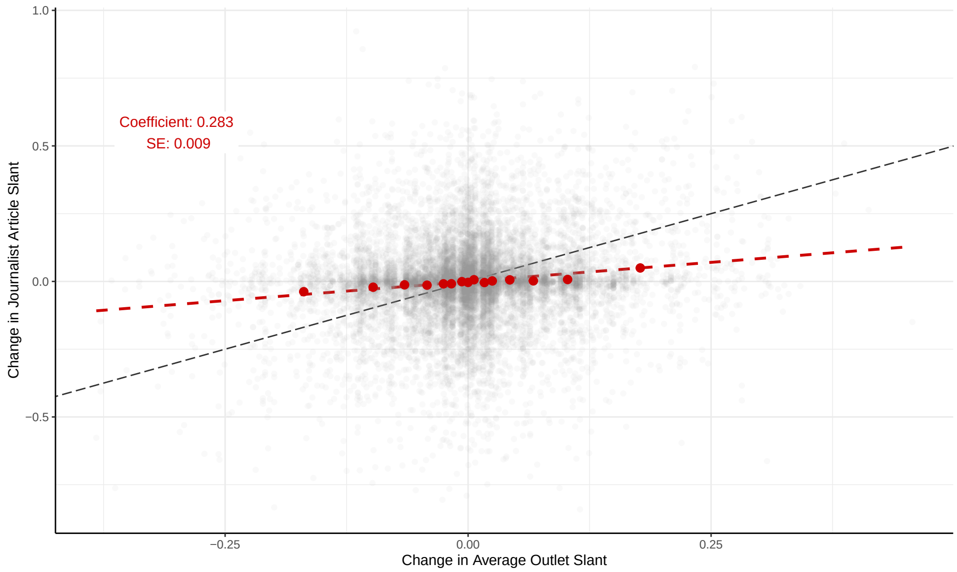
Panel B: Wall Street Journal



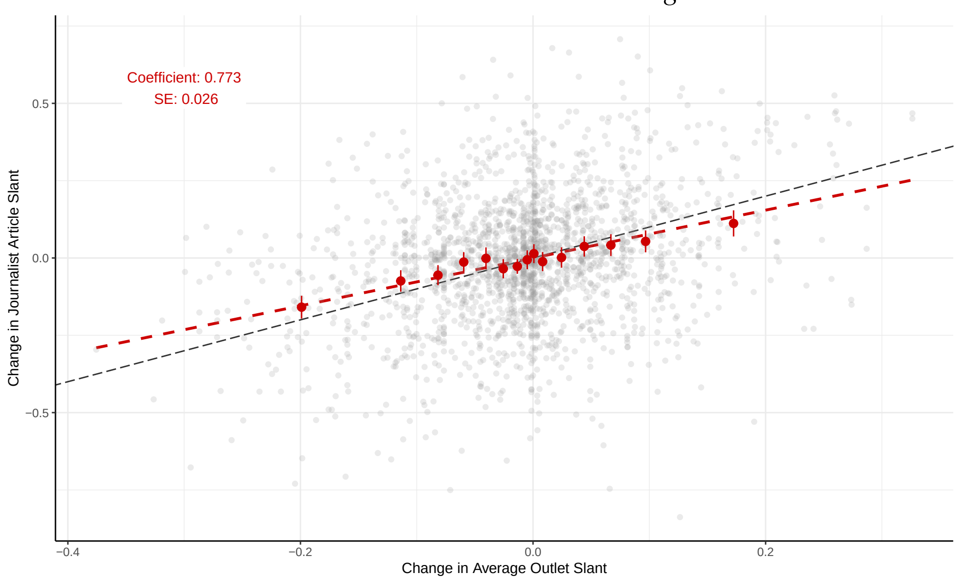
Notes: Figure plots the distribution, across authors, of the average slant of each author's article as measured by our RoBERTa model of slant for each outlet. Panel A includes all authors who appear in *The New York Times* for 2013 and 2018 separately. Panel B is the same as Panel A, except for *The Wall Street Journal*.

Appendix Figure B5: Predicting Change in Journalist Slant—Single Article Change

Panel A: Full Sample

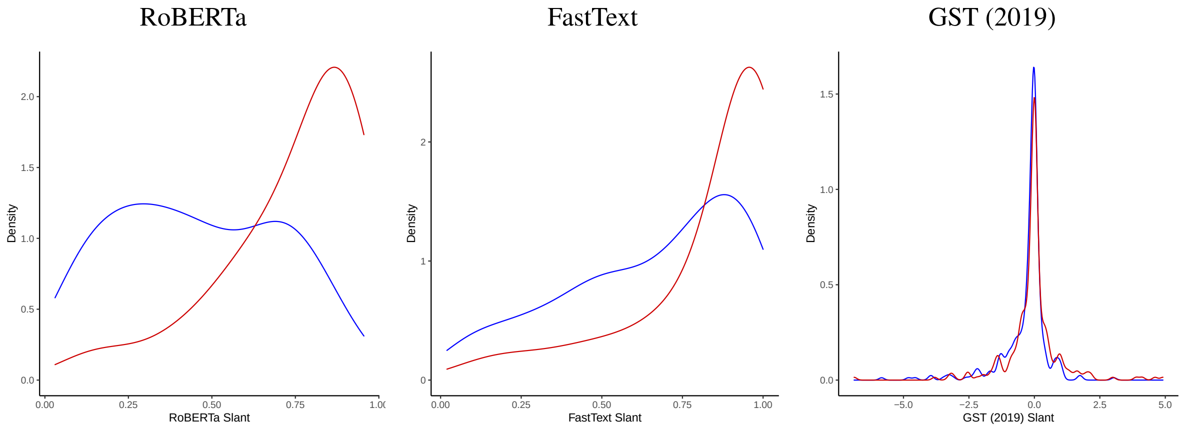


Panel B: More than 5 consecutive articles at origin and destination



Notes: Panel A plots the (bin)scatter plot for our full data sample of movers with $\bar{x}_{d(i)} - \bar{x}_{o(i)}$ on the x-axis and $\hat{x}_{i0} - \hat{x}_{i,-1}$ on the y-axis. Panel B is the same as Panel A, except restricted to authors who wrote more than five consecutive articles at the origin and the destination respectively upon move. Observations are weighted by $(\bar{x}_{d(i)} - \bar{x}_{o(i)})^2$.

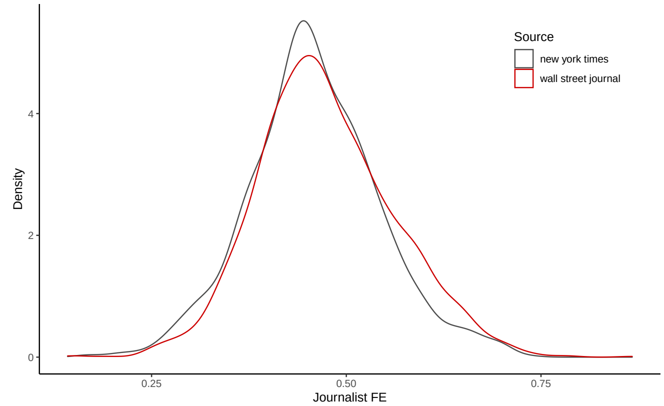
Appendix Figure B6: Comparing to Human Perceived Slant — Distributions



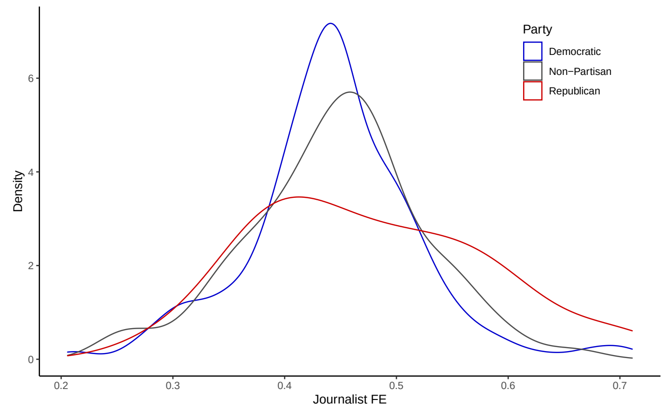
Notes: Figure plots the distribution of the text-based measure separately for articles coded as leaning Republican (red) and articles coded as leaning Democratic (blue) by the MTurk coders in Budak et al. (2016). The sample is restricted to articles for which both MTurk coders gave the same relative rating and for which this relative rating was non-zero.

Appendix Figure B7: Heterogeneity in Journalist AKM Fixed Effects — WSJ and NYT

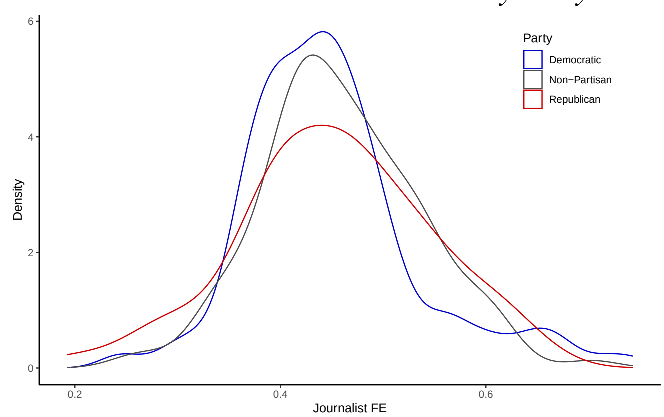
Panel A: New York Times vs. Wall Street Journal



Panel B: New York Times – By Party



Panel C: Wall Street Journal – By Party



Notes: Each panel plots the distribution in estimated AKM journalist fixed effects across various sub-groups of journalists. Panel A examines heterogeneity across journalists at *The New York Times* and *The Wall Street Journal*. Panel B examines heterogeneity at *The New York Times* by partisan affiliation. Panel C examines heterogeneity at *The Wall Street Journal* by party affiliation.

Appendix Table B1: Prediction of Human Perceived Slant from Budak et al. (2016), Article Level

Dependent Variable:	Human Perceived Slant			
	(1)	(2)	(3)	(4)
RoBERTa	0.302	.	.	0.279
	(0.012)	.	.	(0.014)
GST (2019) Dictionary-based	.	0.087	.	0.032
	.	(0.012)	.	(0.012)
FastText	.	.	0.199	0.031
	.	.	(0.012)	(0.014)
Observations	5980	5980	5980	5980
R^2	0.101	0.008	0.044	0.103

Notes: Table shows the results of regressing average human perceived slant as rated by MTurk coders in Budak et al. (2016) on various text-based models of slant. Sample is composed of a random sample of articles from Budak et al. (2016). For some selected articles, we were unable to scrape the underlying text and these articles are removed. In Budak et al. (2016), each article is scored on a 5-point favorability score indicating the extent to which the article is favorable to Republicans and a separate score for how favorable the article is to Democrats. We take the difference between these two scores as our measure of perceived slant. Each article is rated twice—once by a coder who does not observe the outlet's name, and once by a coder who does observe the outlet's name. We average these two scores to obtain our final measure of human perceived slant for each article. The RoBERTa measure is our main measure of slant. The FastText measure is a measure of slant using Facebook's FastText model and trained on a dataset similar to the RoBERTa model. The GST (2019) Dictionary-based measure is computed by taking the top 1000 most partisan phrases of the 114th Congress from Gentzkow et al. (2019) and assigning each article the average score across observed phrases, assuming a partisanship of zero for phrases not appearing in the Gentzkow et al. (2019) list. All text-based model slant estimates are normalized to have a standard deviation of 1.

Appendix Table B2: AKM Decompositions

	(1)	(2)	(3)	(4)
Sample:	Women	Men	White	Non-White
Panel A: Plug-in Estimator				
	Share of Variance, x_{ijt}			
Journalist, var(γ_i)	0.342	0.375	0.344	0.469
Outlet, var(α_j)	0.298	0.244	0.261	0.297
Sorting, $2 \times$ cov(γ_i, α_j)	0.017	0.068	0.074	-0.095
	Correlation			
Sorting, cor(γ_i, α_j)	0.026	0.112	0.124	-0.128
Panel B: Leave-Out Estimator (Kline et al. 2020)				
	Share of Variance, x_{ijt}			
Journalist, var(γ_i)	0.259	0.313	0.287	0.368
Outlet, var(α_j)	0.247	0.217	0.236	0.233
Sorting, $2 \times$ cov(γ_i, α_j)	0.114	0.117	0.121	0.028
	Correlation			
Sorting, cor(γ_i, α_j)	0.226	0.224	0.233	0.049

Notes: Table reports AKM variance decomposition exercises using the plug-in estimator (Panel A) and the Kline et al. (2020) leave-out estimator (Panel B) with various samples. Both panels are restricted to the leave-out connected set of outlets. Columns (1)–(4) restrict to various groups of journalists according to name-based predictions of gender and race.

Appendix Table B3: AKM Decompositions, Full Connected Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Sample:	All	Dropping Stayers (Full)	Dropping Stayers (Large)	12+ Months	Drop 1 Mo. Moves	Top 20 Outlets	$T=5$ Full	$T=5$ Large	$T=10$ Full	Parent Company	Women	Men	White	Non-White
							Share of Variance, x_{ijt}							
Journalist, var(γ_i)	0.353	0.313	0.296	0.367	0.354	0.426	0.257	0.228	0.272	0.527	0.353	0.380	0.346	0.509
Outlet, var(α_j)	0.251	0.257	0.326	0.261	0.256	0.128	0.161	0.187	0.282	0.056	0.304	0.248	0.268	0.322
Time, var(τ_t)	0.006	0.007	0.008	0.006	0.006	0.022	0.000	0.001	0.001	0.008	0.008	0.005	0.005	0.012
Sorting, $2\times$cov(γ_i,α_j)	0.075	0.088	0.041	0.053	0.070	0.006	0.065	0.042	-0.050	0.063	0.006	0.060	0.068	-0.151
Other Components	0.315	0.335	0.329	0.314	0.314	0.419	0.517	0.542	0.495	0.346	0.330	0.307	0.313	0.308
							Correlation							
Sorting, cor(γ_i,α_j)	0.126	0.156	0.067	0.085	0.116	0.013	0.159	0.101	-0.090	0.183	0.009	0.097	0.112	-0.186
							Sample Size							
Observations	701478	167377	44898	657100	694060	252786	70234	60899	16256	585173	257077	428014	561307	61997
Journalists (All)	30048	5513	1498	28807	29585	10114	4281	2411	1094	24895	11949	17224	23728	2781
Journalists (Movers)	5513	5513	1498	4272	4402	579	4281	2411	1094	3503	1913	3483	4461	521
Outlets	298	288	263	298	297	18	284	268	234	37	296	294	291	258
Months	72	72	72	72	72	72	11	21	11	72	72	72	72	72

Notes: Table reports AKM variance decomposition exercises using the plug-in estimator with various samples. Column (1) uses the full sample. Column (2) drops journalists that never switch outlets. Column (3) is the same as Column (2) except that it also restricts to journalists whose range in outlet slant is greater than 0.1. Column (4) drops journalists that appear in more than two outlets. Column (5) restricts to tenures of longer than one month in duration. Column (6) restricts to the top 20 outlets according to weekend circulation. Column (7) uses the $T=5$ event study sample at the article level. Column (8) is the same as Column (7) except that it restricts to moves where $|\bar{x}_{d(i)} - \bar{x}_{o(i)}| > .1$. Column (9) uses the $T=10$ event study sample at the article level. Column (10) treats all outlets owned by a single "Parent Company" as the same outlet and restricts attention to outlets for which we observe the parent company. Columns (11)–(14) restrict to various groups of journalists according to name-based predictions of gender and race.