Does Politicized Media Distort Political Discourse? Evidence from U.S. Cable News Channels

Elliott Ash and Elena Labzina*

July 20, 2019

Preliminary working paper. Comments welcome.

Abstract

While previous work has shown that partisan media affects voter choices, an open question is whether and how partisan news messaging influences the language of political discourse. This paper provides evidence on this influence in the context of the U.S. Congress and major cable news networks for the years 2005 through 2008. We measure media influence using a measure of the similarity between language in Congressional speeches and language used by speakers on Fox News, CNN, and MSNBC. Exogenous variation in news exposure across congressional districts comes from relative channel numbering, which we use as instruments. We find Fox News has had the largest effect on Congressional language, with MSNBC and CNN having little effect. Cable news has no effect on partisanship of roll call votes nor on the partisanship of speech.

*ETH Zurich. Email contacts: ashe@ethz.ch, elabzina@ethz.ch. We thank Greg Martin and Sergio Galletta for help in early stages in the project. We are thankful for comments from audience members at MPSA and especially from our discussant Jason Anastasopoulous. Special thanks to Romina Jafaryanyazdi for helpful research assistance.

1 Introduction

Does partisan media have a causal impact on the discourse of legislators? This is an important question for political scientists and for policymakers making decisions about media regulation. While previous work has shown that partisan media affect voting by citizens and politicians, there is no previous evidence on how media affect choices of language.

This paper seeks to provide evidence on this issue using the full text of the U.S. Congressional Record, matched with the full text of news show transcripts from Fox News, CNN, and MSNBC, for the years 2005 through 2008. These corpora are used to measure differences in the similarity of language to cable news messaging across Congressmen. The empirical analysis is designed to test the hypothesis that in congressional districts which have more exposure to particular news messaging, the language in that messaging would be reflected in the congressional speeches expressed by the associated legislator representing that district. Exogenous variation in news exposure across congressional districts comes from random variation in the relative channel numbering of these three networks.

This is a challenging empirical problem which invites a number of empirical innovations. First, we borrow measures of text similarity from the information extraction literature. The preferred approach is to represent documents in a vector space based on phrase frequencies and then measure the influence of cable news using cosine similarity. This text distance metric assesses the similarity of each congressional speech to the linguistic particularities of the three news networks.

We then ask whether relative similarity to language in a cable news network increases in response to higher Fox News viewership in a Congressman's district. Cross-sectional or panel data estimates of this relationship would likely be biased due to a confounder for conservatism across districts. We procure causal estimates of this effect by using the Fox News channel number as an instrument for Fox News viewership. Because we face a weak instruments problem at the level of congressional districts, we use machine learning methods to extract more predictive power from the distribution of channels in the first stage. We provide a number of checks to validate first stage relevance and exogeneity of the instrument.

Our results can be summarized as follows. Of the three major cable news channels, Fox News has the largest effect on congressional speech. When Fox ratings are higher in a district due to channel positioning, the congressman uses more language associated with Fox News. CNN and MSNBC do not have significant effects. There are no effects on partisanship of speech (as measured by probability of being Republican or Democrat based on speech text features), nor on a predicted partisanship measure constructed from roll call votes.

To understand the mechanism better, we undertook some follow-up analysis. First, we compared the similarity of phrasing versus similarity of topics, and the evidence suggests Fox distorts discourse through how topics are framed, rather than choice of topics covered. Second, we don't see immediate responses to weekly changes in Fox language, suggesting that the results are driven by durable shifts in how constituents respond to political language, rather than by active pandering to the short-run interests of Fox viewers.

This research combines methods from natural language processing, machine learning, and causal inference, adding to the literature on the use of text data to understand ideological influences in politics. This literature includes Gentzkow and Shapiro (2010), Gentzkow et al. (2017), and Ash et al. (2017), who analyze divisiveness in congressional language. Those papers focus on what phrases are associated with what party using a supervised approach. Our similarity-based methods and instrumental-variables methods are useful for computational social scientists seeking to use text in a causal framework. As detailed below, we address a number of issues in terms of high dimensionality, computational intractability, lack of interpretability, and omitted variable bias.

The substantive results add to the literature in political science and political economy on the role of media in electoral politics (e.g. Ashworth and Shotts, 2010; Prat, 2017). In particular, we look at the effect of media on the behavior of U.S. Congressmen. A key paper is Snyder and Stromberg (2010), who show that higher media coverage due to random variation in overlap between newspaper markets and congressional districts increased effort by congress members on a range of measures. Another strain of papers focused on the influence of cable news in politics. One set of papers has shown they have impact on voter knowledge about political issues (Hopkins et al., 2014; Schroeder and Stone, 2015). In addition, there is now good evidence that Fox News (in particular) has an impact on vote shares in presidential elections. The closest papers to this one have shown that there is a correlation between Fox viewership and politician position-taking (Clinton and Enamorado, 2014; Arceneaux et al., 2016), although those results are sensitive to context.

More broadly, the study contributes to the long-lasting debate on the importance of (un)biased media in democratic politics. This topic has become especially important in the current era of extreme polarization and increasing inequality in American society. If the biased media has explicit effects on the discourse in Congress, it will likely affect the enacted policies.

2 Data

This section enumerates our data sources. The data come from cable news channels and from U.S. Congress. Our resulting panel is from 2005 through 2008 because those are the years for which we could construct cable news viewership by congressional district.

First, we have data on channel positions and ratings. These are from Nielsen and are the same as the data used by Martin and Yurukoglu (2017). The original data are at the zip code level. For our analysis, these are aggregated to the congressional district level, using the population-weighted zip code averages. This is done both for the ratings and the channel positions. We have information for Fox, CNN, and MSNBC. The baseline viewership variable, ratings, is proportional to the number of minutes spent watching a channel on average per household. The results are robust to instead using viewer share, which gives the proportion of television time devoted to the corresponding channel.

Second, we have collected the corpus of news show transcripts for Fox, CNN, and MSNBC.

These are collected for all prime time shows from LexisNexis. We have a series of scripts that reads through the transcripts and excludes metadata and other non-speech content.

Third, we have the full text of speeches from U.S. Congress floor debates. These are from the Congressional Record, used previously in Ash et al. (2017). Again, we have a set of scripts that reads through the speeches and extracts the speaker and the date, along with the plain text.

Fourth, we have a range of metadata on Congressmen. From *www.congress.gov* we have personal characteristics such as district, party, and gender. From *voteview.com* we have DW-NOMINATE scores, which are a standard metric for ideology in roll call votes (Poole and Rosenthal, 2001). Briefly, DW-NOMINATE is the first principal component of the matrix of roll call votes, normalized such that a higher number is interpreted as ideologically conservative and a lower number is ideologically liberal.

Finally, we have a rich set of demographic covariates from the 2000 census. These are averaged by zip code, weighted by zip code population, to get the aggregate value for the congressional district. Summary statistics for these variables are reported in Table C.7. For the regressions, including all these variables as covariates was not advisable due to high collinearity in some variables. Therefore we took principal components for the matrix of demographic covariates. The Scree plot for these components is shown in Figure C.1. Based on this Scree plot, we use the first eight principal component scores as covariates in our regressions.

3 Text Data Methods

This section describes how we construct language measures from the text of speeches and news transcripts.

3.1 Quantification of the text data

We start with the plain texts of the congressional speeches and TV transcripts. First, we do the preprocessing of the texts, which includes converting them to lower case and removal of non-meaningful words, all non-letter characters, and extra white-spaces. Second, for each word we perform stemming, removing stems to reduce each word to its root. Now, for each observation instead of having plain text, we have a list of meaningful stemmed tokens.

Our aim at this stage is to transform the texts to numerical data of a minimal relevant dimensionality. However, since we would like to be able to explicitly track the effect of distinct stems, we opt out of employing the LSA approach (Pincombe, 2004). The first obvious solution, classical "bag-of-words" representation, does not capture the sequential structure of a text. One way to extend the bag-of-words approach to reflect word order is to use n-grams. N-grams are sequential groups of tokens constructed based on the initial list of tokens. For example, if we construct 3-grams from the list such as (stem1, stem2, stem3, stem4) then it converts to the list of two 3-grams, ((stem1, stem2, stem3), (stem2, stem3, stem4)).

Then, we create the dictionary of features for each of the channels and the Congress speeches. By going through all observations, we build the frequency distribution over n-grams in our data and assign a position number to each n-gram based on frequency. For our main specification, we select the central 95% of the 3-grams frequency distribution, removing the least frequent and most-frequent n-grams from the dictionary. In unreported results, we found qualitatively similar but noisier results when using the central 99% of the frequency distribution. We select the former specification because it provides a nice trade-off between being parsimonious (not having too many features) and still capturing the semantic similarity.

Let V give the set of phrases, indexed by i, from this final vocabulary included in the central 95% of the frequency distribution. Let A_{it}^{j} be the frequency of phrase $i \in V$ for congressman j at time t; specifically, the number of times i appears in floor speeches by j

given during t. Let B_{it}^{k} be the frequency of phrase i for cable channel k at time t; specifically, the number of times i appears in a transcript for shows on k during t. In terms of time periods, we compute frequencies by month and by year. In vector notation, we have $\mathbf{A}_t^j$ and $\mathbf{B}_t^k$ giving the vectors of frequencies across the vocabulary of phrases, for congressmen and cable channels respectively.

How do Fox, CNN, and MSNBC differ in their use of language? To understand this, we took the set of trigrams in our vocabulary and computed the correlation of the relative frequency in each network. We then took the top ten trigrams most associated with each network. These are listed in Table C.8. Unlike the distinctively partisan phrases between political parties demonstrated in previous work (e.g. Gentzkow and Shapiro, 2011), the distinguishing phrases between the cable news channels are not particularly illuminating of the ideologies of the networks.

3.2 Congressional speech similarity to cable news

The next step is to estimate the similarity between the TV transcripts and congressional speeches. We want to estimate the distance, or similarity score, for each House Member, and for each TV channel, and for each time period (year or month).

There are multiple methods to calculate distances between two vectorized documents. To begin, we will use the "geometrical" approach: *cosine similarity*. Cosine similarity is the standard document comparison metric in the fields of text classification and information extraction (e.g. Li et al., 2003). Other well-known "similarity" methods include probabilistic ones (for instance, based on LDA (*Hellinger distance*)) and based on set theory principles (*Jaccard distance*) (Huang, 2008). A feature of our implementation of the cosine similarity is the use of a bag-of-ngrams representation of documents, rather than bag-of-words format.

The idea behind the cosine similarity is simple: once each observation becomes a non-negative vector, the similarity for each two vectors can be estimated as the angle between them. In linear algebra terms, it is the dot product between two normalized vectors.

The cosine measure is monotonically decreasing from 1 to 0 on $[0, \pi/2]$:

$$\text{sim}(obs_a = A, obs_b = B) = cos(AB) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{1}$$

From (1), it is easy to notice that the metric does not depend on the magnitudes of A and B, only on the angle between the *rays* they represent. This means that it is not sensitive to document length.

The pros as well as the cons of the cosine similarity method come from its simplicity. It is transparent and fast to compute. On the other hand, it requires a pre-defined vocabulary, and treats all text features as independent. So for example the metric doesn't "know" that "taxes" and "revenues" are synonymous, or that a document mentioning "taxes" is probably related to "budgets". More recent text similarity metrics based on word embeddings have tried to address this issue (Kusner et al., 2015), with the disadvantage that they don't capture local word order from N-grams and generally have higher computation costs.

In our setting, we are interested in the textual similarity of each congressman j to each news channel k in a given year t. Formally, we compute

$$Y_{jt}^{k} = \text{sim}(\mathbf{A}_t^j, \mathbf{B}_t^k), \forall j, k \tag{2}$$

and can use this as a baseline similarity measure. Our preferred specification is *relative similarity*, where the similarity to Fox, for example, is divided by the average similarities to CNN and MSNBC, and then standardized to variance one. Analogous measures are computed for CNN and MSNBC. Results are robust to standardizing the raw similarities before computing the relative measure. Appendix A discussed a range of robustness checks for computing the similarity.

The cosine similarity approach diverges from most of the previous literature on political influences in language, which takes a supervised approach (e.g. Gentzkow et al., 2017). In brief, these papers take the party of a speaker as a label to be predicted based on the speaker's language. Then a measure of speech partisanship would be the share of language used that is predictive of party affiliation. We produce this type of measure as an alternative outcome for our analysis, as described in the next subsection. The partisanship measure is useful for answering the interesting research question on whether cable news media cause Democrats or Republicans in Congress to become more partisan in their language. But our main question in this paper is somewhat broader, on whether news media language generally gets into politics (and not just that associated with the parties). Cosine similarity provides a cleaner measure of this phenomenon.

3.3 Partisanship of Speech Text

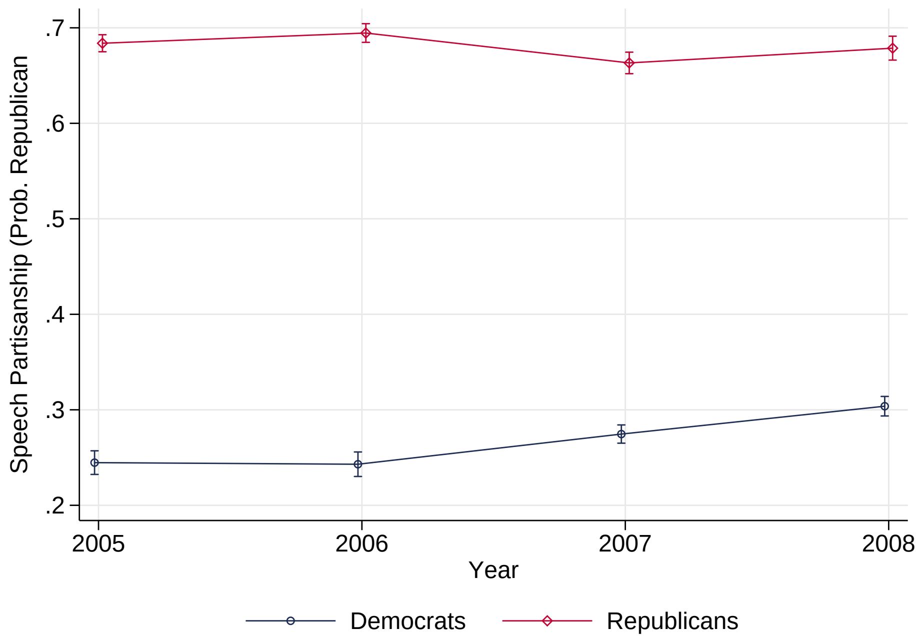
As mentioned, most of the previous work has taken a supervised learning approach to predict party from text. If party can be predicted more accurately by the text, then that means the text is more polarized. In turn, one can compare speakers, and say their text is more partisan than others, by forming the party label predictions from the associated text features.

Our research question is whether the language used by news media causally distorts the language of politicians. A more targeted question is whether the media have an influence in a particular partisan direction. To get at this question, we follow an approach similar to Gentzkow et al. (2017) to score the partisanship of speeches. We then use that as an outcome in our cable channel regressions.

For this task, we started with congressional speeches between 1994 and 2015, with speeches aggregated at the monthly level. The speech tokens were filtered to include nouns, verbs, and adjectives, and then bigrams were constructed from those filtered tokens. After taking the 10,000 most frequent bigrams as the full feature set, we then used chi-squared feature selection (trained on speaker party) to select 2,000 predictive bigrams for the machine learning feature set.

These features were then used in an L2-regularized logistic regression model to predict speaker party. We used grid search to pick hyperparameters (regularization = 4) and then

Figure 1: Predicted Republican Scores By Year and Party



scored the resulting model using cross-validation. The model predicts party far better than chance, with mean accuracy .76 (std. dev. .025) across five folds.

As a descriptive exercise, we looked at the partisanship of language in the cable news transcripts. On this metric, there is a big difference between Fox, on the one hand, and CNN/MSNBC, on the other hand. Based on the model, both CNN and MSNBC segments were predicted to be Democrat over 90 percent of the time. In contrast, Fox transcript segments was predicted to be Democrat only 50 percent of the time.

For the empirical analysis, we define $R_{jt} = \Pr(\text{Republican}|A_{jt})$, where A_{jt} gives the speech text for congressman j at time t. The values for this variable, separately by congressman and year, are plotted in Figure 1. In our analysis, this will be used first as a control in our cosine-similarity analysis to test the robustness of our effect estimate. Second, it is an alternative outcome for the effect of cable news.

4 Econometrics

The data on text similarities are combined with the data on news viewership for the empirical analysis. The main hypothesis is that higher news channel viewership by constituents will cause a congressional representative to address similar issues to those discussed on that news channel. This section outlines our method for testing this hypothesis.

4.1 Linear Regression Framework

For the purposes of illustration, let Y_{it} give the relative similarity to Fox News for congressman i at month t (see subsection 3.2). In turn, the treatment variable on viewership, X_{it}, is the constituency-level viewership for Fox News (from Martin and Yurukoglu, 2017), aggregated across zip codes (weighted by population).

Now our problem can be formulated as a standard linear regression:

$$Y_{it} = \alpha_{it} + \theta X_{it} + \epsilon_{it}, \tag{3}$$

where θ is our effect of interest; α_{it} can include state-time fixed effects and controls, and the error term includes omitted variables and randomness. Eq. (3) could then be estimated by ordinary least squares (OLS) .

The problem with OLS estimation of (3) is the potential correlation between X_{it} and ϵ_{it}. There are many political and economic factors that may be correlated with both Fox News viewership and a Congressman's use of Fox News language, in particular the pre-existing ideological preferences of the district. This endogeneity will lead to bias in the estimate for θ.

To address this problem, we take an instrumental variables approach based on Martin and Yurukoglu (2017). We require that the instrument Z_{it} is correlated with X_{it}, but not confounded with other factors affecting Y_{it}. The following subsections describe how we construct instruments and then estimate two-stage least squares.

4.2 Zero-stage cross-validated ridge regression

Martin and Yurukoglu (2017) instrument for cable news ratings using the cable channel positions (that is, the number in the TV channel lineup) across zip codes. In our case, we would average across zipcodes to compute congressional-district-level data with population weighting. We have found, however, that we cannot replicate the strong first stage from Martin and Yurukoglu (2017) at the level of the congressional district, probably because we have only 435 clusters (compared to the 4,700 clusters in Martin-Yurukoglu). Therefore following their method exactly is likely to have a large bias.

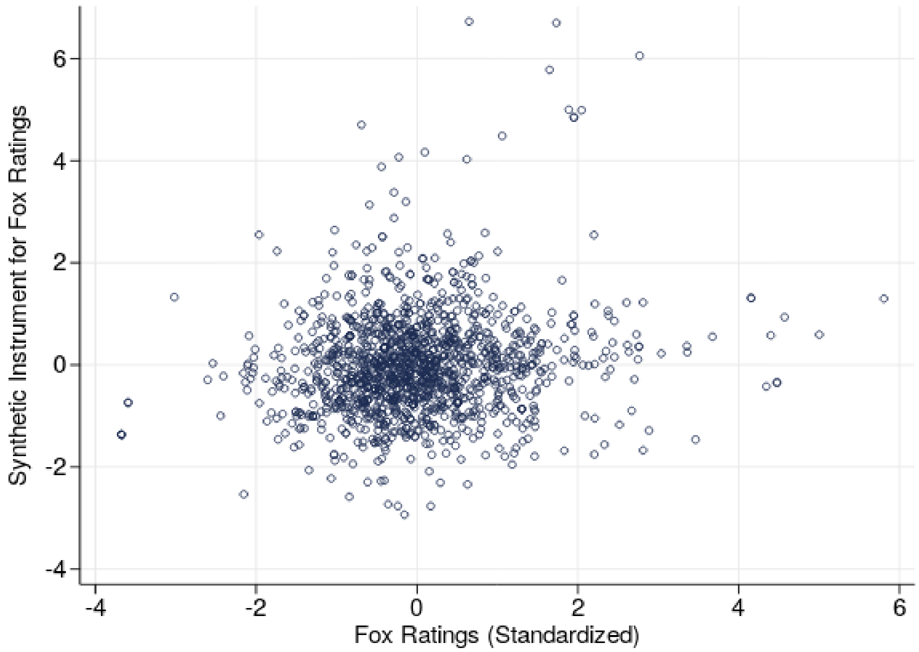
We address the weak-instruments issue by extracting more predictive first-stage information from the distribution of channels. Our "zero-stage cross-validated ridge regression" draws on recent methods in high-dimensional econometrics (Belloni et al., 2012; Hansen and Kozbur, 2014; Hartford et al., 2017; Chernozhukov et al., 2018). In particular, the approach has the statistical properties of regularized jack-knife IV but without the computational costs of computing a first-stage coefficient for each observation separately (Hansen and Kozbur, 2014).

The method works as follows. First, the channel positions for each network are residualized on state-year fixed effects and the demographic covariates (principal components). Then the residuals are normalized to variance one.[1] The endogenous regressor, cable news ratings, is also residualized on state-year fixed effects and the covariates. Next, the full set of interactions and quadratic transformations are formed from the channel positions. We then use them as predictors in a ridge regression to predict residualized ratings. We learn the L2 penalty using 10-fold cross-validation grid search; in the case of Fox ratings, for example, the average penalty across folds is 0.1425.

The ridge regression coefficients are used to form cross-validated predictions for the endogenous regressor using the original instruments. This prediction $Z_{it} = \hat{X}_{it}$ will then be

[1] This step is standard in regularized regression, so that penalized coefficients are comparable (e.g. Hastie et al., 2009).

Figure 2: Scatter Plot of Fox Ratings vs. ML Instrument



used as the actual univariate instrument for the corresponding news network ratings. It is a "clean" prediction in the sense that the coefficients are trained on out-of-fold data. Analogously, jack-knife instruments gain independence from using a leave-one-out fitted value (Angrist et al., 1999; Hansen and Kozbur, 2014); the only difference is that 90% of the dataset is used for fitting the instrument, rather than $N-1$. The instrument is strongly correlated with the ratings, but far from collinear, as shown in Figure 2's scatter plot.[2]

[2]Why is this Z_{it} a completely different variable in relation to X_{it}? The explanation why $\hat{X}_{it}$ is a separate variable and can serve as an instrument for X_{it} comes from the following argument. If we think that the set of X_{is}'s where s $\in \{1, n\}$ is conditionally independent, then the model, $\hat{X}()$, trained on its subset for $s \in \{1, m\}$ such that $t \notin \{1, m\}$ does not have any formal or endogenous relation to X_{it}. So neither is the realization of the model, $\hat{X}_{it} = \hat{X}(X_{it})$. Indeed, as our tests show, $X_{i,t}$ is not endogenous with $\hat{X}_{it}$ but is strictly correlated with it, which is logical since $\hat{X}(X_{it})$ approximates X_{it} based on the existing information. On top of that, by construction, it passes the exclusion assumption. Hence, $Z_{it} = \hat{X}_{it}$ can serve as an instrument for X_{it}.

Table 1: First Stage Estimates: Unadjusted Channel Positions

	(1) FNC Ratings	(2) CNN Ratings	(3) MSNBC Ratings
FNC Channel Position	-0.0818*		
	(0.0420)		
CNN Channel Position		-0.152***	
		(0.0442)	
MSNBC Channel Position			-0.106***
			(0.0370)
Observations	1,398	1,392	1,392
R-squared	0.007	0.023	0.011
F-test	3.799	11.91	8.232

Regressions include state-year FEs. SEs in parentheses clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

4.3 First Stage

Now we have the first stage regression

$$X_{it} = \alpha_{it} + \gamma Z_{it} + \eta_{it}, \tag{4}$$

which, combined with (3), can be estimated with two-stage least squares (2SLS) to procure consistent estimates for θ. The preferred specification includes state-time fixed effects and demographic controls for the year 2000. For inference we cluster standard errors by congressional district, to allow for serial correlation over time within district. If cable news affects congressional speeches through the constituency, then we would estimate $\theta > 0$.

2SLS requires relevance in the first stage. We check for relevance using the robust first-stage F-statistic from Stock et al. (2002) for the case of one instrumental variable. As visualized in Figure 5 and reported statistically in Table 2, our ML-predicted instruments provide sufficient power in the first stage. In the tables below, we also report Kleinbergen-Paap cluster-robust first-stage F-statistics.

Figure 3: The first stage: FNC

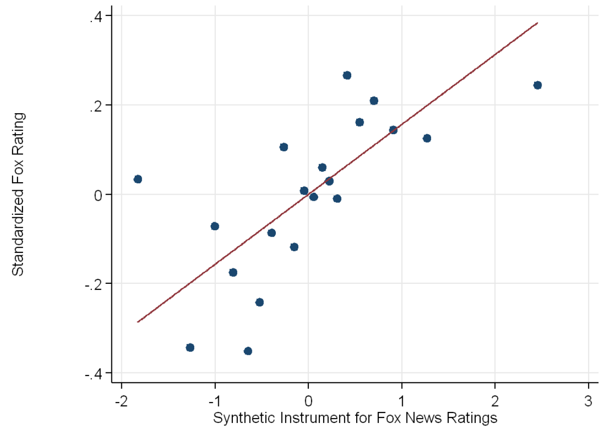


Figure 4: The first stage: CNN

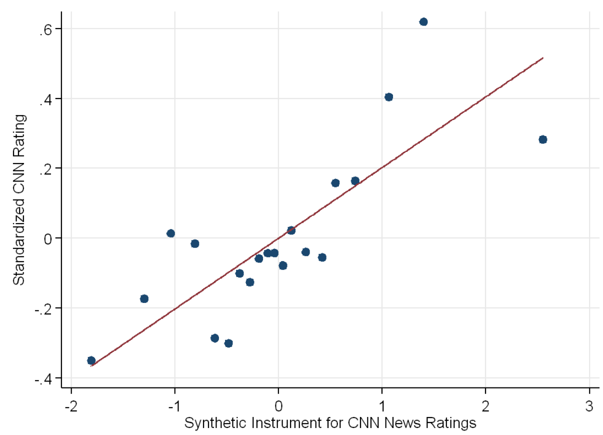


Figure 5: The first stage: MSBNC

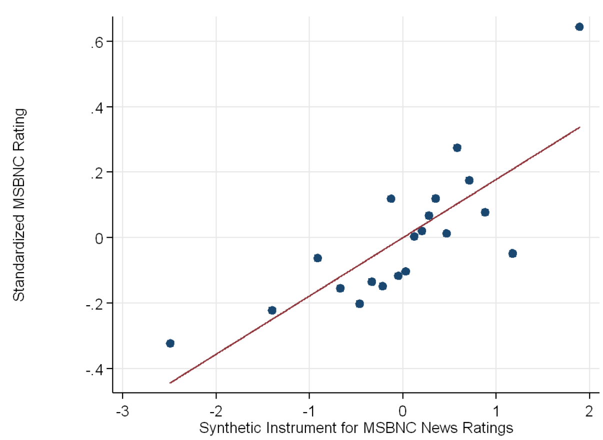


Table 2: First Stage Estimates: ML Instruments

	(1) FNC Ratings	(2) CNN Ratings	(3) MSNBC Ratings
Z_fnc	0.154***		
	(0.0412)		
Z_cnn		0.201***	
		(0.0484)	
Z_msnbc			0.176***
			(0.0607)
Observations	1,398	1,392	1,392
R-squared	0.023	0.039	0.028
F-test	14.05	17.19	8.378

Regressions include state-year FEs. SEs in parentheses clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

4.4 Identification Checks

Next, we need exogeneity and exclusion. The latter (the exclusion restriction) is not controversial in our case, as channel positions are unlikely to have any independent effects on political activities besides their effects on channel viewership. The major concern is exogeneity, which requires that our channel position instrument be orthogonal to any other factors that affect our outcome. Put differently, there are no unobserved political, economic, or cultural factors across congressional districts that would affect both the channel number and the similarity metric.

Martin and Yurukoglu (2017) provide a lengthy discussion and set of checks for exogeneity. Drawing on historical sources, they argue that channel positions have an important arbitrary / historical component with significant inertia and path dependence. This means that networks play only a limited role in influencing the network's position and cannot easily adapt it to local conditions. Once the channel is set early on, it tends to stick around the same channel number as not to confuse viewers. Possible minor alterations are due to the changes of the set of the available channels over time.

In our data, we undertake a number of identification checks. As in Martin-Yurukoglu, the

Table 3: Channel Positions are unrelated to Past Republican Vote

	(1)	(2)	(3)	(4)	(5)	(6)
	Effect on 1996 Republican Presidential Vote Share					
FNC Channel	0.000375					
	(0.000558)					
Z_fnc		-0.00321				
		(0.00416)				
CNN Channel			8.78e-05			
			(0.000668)			
Z_cnn				-0.000871		
				(0.00460)		
MSNBC Channel					0.000468	
					(0.000345)	
Z_msnbc						-0.00392
						(0.00362)
Observations	1,398	1,398	1,398	1,398	1,398	1,398
R-squared	0.531	0.532	0.530	0.530	0.533	0.532
F-test	0.451	0.597	0.0173	0.0358	1.840	1.175

Regressions include state-year FEs. SEs in parentheses clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Channel Positions Unrelated to Similarity of 1995 Speeches to Current Transcripts

	(1)	(2)	(3)	(4)	(5)	(6)
	1995 Speech Sim(Fox)		1995 Speech Sim(CNN)		1995 Speech Sim(MSNBC)	
Z_fnc	0.0106	0.0314				
	(0.0149)	(0.101)				
Z_cnn			-0.00273	-0.0149		
			(0.0133)	(0.0817)		
Z_msnbc					0.00629	0.0160
					(0.0107)	(0.0740)
Observations	756	756	756	756	756	756
R-squared	0.185	0.155	0.167	0.151	0.129	0.138
F-test	0.508	0.0969	0.0424	0.0332	0.345	0.0467

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 5: Channel Positions are Unrelated to Predictive Demographics

	(1) Predicted Sim(Fox)	(2) Predicted Sim(CNN)	(3) Predicted Sim(MSNBC)
Z_fnc	-0.000795		
	(0.00222)		
Z_cnn		0.0127	
		(0.00865)	
Z_msnbc			0.00278
			(0.0121)
Observations	1,398	1,398	1,398
R-squared	0.555	0.373	0.273
F-test	0.128	2.156	0.0527

Standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

channel positions are not related to pre-treatment (1996) Democrat/Republican vote shares (Table 3). As an additional placebo check, we computed the similarity of pre-treatment congressional speeches (1995) to current transcripts (2005), and regressed those on the 2005 channel positions. There was no effect (Table 4). Third, we regressed the 2005-2008 similarity metrics based on 2000 demographic variables, and then used the predicted $\hat{Y}_{it}$ in the reduced form. Again, there was no effect (Table 5).

5 Results

This section reports the results.

5.1 OLS and Reduced Form

Let's start with ordinary least squares estimates. Table 6 reports these estimates, which give the cross-sectional relationship between news channel ratings in a district and the representative's similarity of language to that network's messaging. There is a positive and significant relationship for FNC and MSNBC.

Table 6: OLS Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sim to Fox		Sim to CNN		Sim to MSNBC	
Fox Ratings	0.0473**	0.0476*				
	(0.0235)	(0.0251)				
CNN Ratings			0.0162	0.0232		
			(0.0199)	(0.0195)		
MSNBC Ratings					0.0400**	0.0346*
					(0.0193)	(0.0193)
Demo Controls		X		X		X
Observations	1,398	1,398	1,392	1,392	1,392	1,392
R-squared	0.398	0.414	0.256	0.266	0.346	0.358
Adj. R^2	0.322	0.335	0.164	0.170	0.264	0.274

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Reduced Form Estimates

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Sim to Fox		Sim to CNN		Sim to MSNBC	
Z_fnc	0.0532**	0.0578**				
	(0.0243)	(0.0256)				
Z_cnn			0.0130	0.0180		
			(0.0207)	(0.0214)		
Z_msnbc					0.0149	0.0127
					(0.0281)	(0.0273)
Demo Controls		X		X		X
Observations	1,398	1,398	1,398	1,398	1,398	1,398
Adj. R^2	0.322	0.336	0.300	0.302	0.266	0.276

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Next we report the reduced form results. Table 7 shows that the Fox instrument has a positive and significant effect on speech similarity to Fox. For CNN and MSNBC, the reduced-form effect is not significant. The binscatter diagrams in Appendix Figure C.2 add visual evidence by comparing the reduced form relationship to the placebo reduced form, where the 2005 instrument is related to the similarity of 1995 speeches to 2005 transcripts. We can see that all the placebos (right side) are zero. On the left side, only Fox (top panel) has a strong relationship.

5.2 Main 2SLS Results

Now let's look at the results for 2SLS. As Table 8 presents, the effect of the Fox News Channel ratings is positive and significant (Column 1). The effect is robust to inclusion of district demographic controls (Column 2) and to including political party (Column 3). The coefficient is similar, yet more noisily estimated, when including the text-predicted political party (Column 4). As expected, Republicans are more similar to Fox in their language.

In contrast, there are no effects of instrumented ratings on similarity to the network for CNN (Table 9) or for MSNBC (Table 10). Intuitively, we see that Democrats are more similar to MSNBC.

Appendix A provides a lengthy set of robustness checks based on how the vocabulary and similarity score were constructed. The main result – that Fox News exposure of constituents increases a Congressman's speech similarity to Fox news language – is robust to broad changes in the metric. In particular, we tried six different text-distance weighting functions from the literature and obtained similar results. We tried different vocabulary sizes, as well as randomly dropping sets of features, and also got broadly similar results.

5.3 Topics vs. Framing

To understand our effect in more detail, in this section we unpack the results using a topic model. We ran LDA (latent dirichlet allocation) on our speeches using 32 topics (descriptions

Table 8: 2SLS Estimates: Fox

	(1)	(2)	(3)	(4)
	2SLS Effect on Sim to Fox			
Fox Ratings	0.537**	0.658**	0.611**	0.703*
	(0.233)	(0.308)	(0.305)	(0.390)
Republican			0.202***	
			(0.0734)	
Prob(Repub|Text)				0.269
				(0.181)
Demo Controls		X		
Observations	1,398	1,398	1,398	1,326
Adj. R^2	-0.309	-0.403		
Kleibergen-Paap F	11.75	8.827	11.28	8.342

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 9: 2SLS Estimates: CNN

	(1)	(2)	(3)	(4)
	2SLS Effect on Sim to CNN			
CNN Ratings	0.0519	0.0749	0.0521	0.0736
	(0.103)	(0.0982)	(0.102)	(0.114)
Republican			0.00694	
			(0.0562)	
Prob(Repub|Text)				0.123
				(0.115)
Demo Controls		X		
Observations	1,392	1,392	1,392	1,320
Adj. R^2	-0.128	-0.123	-0.129	-0.132
Kleibergen-Paap F	19.10	22.08	19.05	17.24

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 10: 2SLS Estimates: MSNBC

	(1)	(2)	(3)	(4)
	2SLS Effect on Sim to MSNBC			
MSNBC Ratings	0.0811	0.0745	0.0970	0.0415
	(0.161)	(0.173)	(0.155)	(0.160)
Republican			-0.179***	
			(0.0650)	
Prob(Repub|Text)				-0.433***
				(0.139)
Demo Controls		X		
Observations	1,392	1,392	1,392	1,320
Kleibergen-Paap F	8.432	7.772	8.724	7.857

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

available upon request). We will now use these topics to try to understand whether the Fox effect on discourse is due to the choice of topics, or to how topics are framed, or both.

To get at this issue econometrically, we first re-run the 2SLS regressions after including the full set of topic shares by Congressmen as controls. If we still see an effect, that means there is a within-topic framing effect of Fox News. Second, we compute the cosine similarity of topic shares for a Congressman in a year to the topic shares to each network. If this is positive, then there is a change in topic choice due to cable news exposure.

The results of the topics analysis are reported in Table 11. First, we see in Column 1 that our main 2SLS estimates are robust to the inclusion of the topic share controls. Therefore, Fox has an effect on the language in how topics are framed. As before, there is no effect for CNN or MSNBC. Next, in Columns 4 through 6 we see that there is no effect of cable news exposure on the similarity of topic choices to the three networks.

These results support the view that Fox News exposure changes political discourse in the phrase choices used to frame topics. There is not a significant effect on what topics are chosen to discuss. These results are useful in light of the current popularity of topic models

Table 11: Topics or Framing?

	2SLS Effect on Sim			2SLS Effect on Topic Sim		
Fox	0.706**			0.00214		
	(0.338)			(0.113)		
CNN		0.0818			-0.0748	
		(0.127)			(0.120)	
MSNBC			0.201			0.184
			(0.195)			(0.146)
Topic Share Controls	X	X	X			
Observations	1274	1268	1268	1047	905	966
Cragg-Donald F	11.15	49.33	21.81	12.33	28.24	16.18
Kleinbergen-Paap F	6.441	16.70	9.681	6.770	14.00	7.516

Standard errors in parentheses, clustered by District. Demo controls included.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

across the social sciences.

5.4 Is Fox Just Supporting Republicans?

An important question is whether similarity to Fox News is actually a proxy for support for Republican policy positions. To get at this issue we analyze, 1) the probability of getting a Republican Congressman, 2) support for Republican bills (as measured by DW-NOMINATE), and 3) the text-measured partisanship of speech. The latter outcome, as described above, is defined as the probability of being Republican based on text features.

Table 12 looks at the effect of Fox exposure on DW-NOMINATE Scores. We can see that there is no effect of cable news ratings. Table 13 looks at speech partisanship. Again, there is no effect of cable news exposure on these outcomes. In unreported results, we also used party of the congressman as an outcome and saw no effect. Overall, these results suggest that our effect is not working through changing support for partisan priorities.

Table 12: Effects of Cable News on DW-NOMINATE (Republican Roll Call Voting)

	Effect on DW-NOMINATE SCORE					
Fox Ratings	0.321	0.00700				
	(0.345)	(0.127)				
CNN Ratings			-0.233	-0.175*		
			(0.290)	(0.0737)		
MSNBC Ratings					0.684	0.114
					(0.404)	(0.0852)
Republican		1.776***		1.764***		1.780***
		(0.0446)		(0.0452)		(0.0425)
N	2035	2035	2025	2025	2025	2025

Standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Effects on Speech Partisanship

	(1)	(2)	(3)	(4)	(5)	(6)
	2SLS Effect on Speech Partisanship, Prob(Repub|Text)					
Fox Ratings	0.106	0.00257				
	(0.0881)	(0.0585)				
CNN Ratings			-0.0286	-0.0189		
			(0.0739)	(0.0294)		
MSNBC Ratings					0.0166	-0.0375
					(0.0741)	(0.0380)
Republican		0.399***		0.400***		0.399***
		(0.0210)		(0.0162)		(0.0158)
Observations	1,326	1,326	1,320	1,320	1,320	1,320
Adj. R^2	-0.259	0.637	-0.146	0.628	-0.145	0.613
Kleibergen-Paap F	8.864	5.534	17.30	18.79	7.721	7.984

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

5.5 Dynamics of the Effect

In unreported results available from the authors, we looked at the short-run dynamic influence of Fox News on congressional language choices. To this end, we constructed a speaker-phrase-month dataset, and looked at whether deviations in Fox use of a phrase (relative to the other networks) was associated with deviations in speaker use of the phrase. We interacted this effect with having high ratings for Fox, or having a low Fox channel position. We did not find an effect in either case.

Therefore we can rule out a short-term effect of Fox on congressional language, at least as mediated through constituent exposure. This means that congressmen (or their aides) are not apparently attending to and responding directly to the news networks frames in order to pander to constituents. Instead, it is a long-run effect. This means that the congressmen are responding to durable changes in constituent priorities. Fox is changing the style of language or rhetoric that the constituents respond to, and congressmen are following suit.

6 Conclusion

This paper has analyzed the causal effect of media messaging on political discourse. We analyze an important context: how cable news in the United States impacts the language of legislators in the U.S. Congress. Our analysis shows that in response to exogenous variation in viewership, the language of Congressmen is more likely to echo the language of Fox News. The effect is driven by framing of topics, is larger for younger Congressmen, and consists of durable changes in rhetorical style.

The approach to constructing a zero-stage instrument may be unfamiliar because it mixes two statistical paradigms. Ridge regression and cross-validation come from machine learning, which is more modern and has a focus on prediction. But instruments (and specifically jackknife IV) come from econometrics, and classical statistics focusing on inference. Actually, we are mixing three epochs of statistics, according to Efron and Hastie (2016). We do classical inferential statistics when we estimate our 2SLS regression, we do early-age computational statistics with the ridge regression, and, finally, we employ modern-era prediction-style statistics when we utilize the predictions through a cross-validation-style approach. We hope this mixing of approaches is useful for the emerging literature on computational analysis of text (Lucas et al., 2015).

References

Angrist, J. D., Imbens, G. W., and Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics*, 14(1):57–67.

Arceneaux, K., Johnson, M., Lindstädt, R., and Wielen, R. J. V. (2016). The Influence of News Media on Political Elites: Investigating Strategic Responsiveness in Congress. *American Journal of Political Science*.

Ash, E., Morelli, M., and Van Weelden, R. (2017). Elections and divisiveness: Theory and evidence. *Journal of Politics*.

Ashworth, S. and Shotts, K. W. (2010). Does informative media commentary reduce politicians' incentives to pander? *Journal of Public Economics*, 94(11-12):838–847.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Clinton, J. D. and Enamorado, T. (2014). The national news media's effect on congress: How fox news affected elites in congress. *Journal of Politics*, 76(4):928–943.

DellaVigna, S. and Kaplan, E. (2007). The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Efron, B. and Hastie, T. (2016). *Computer age statistical inference*, volume 5. Cambridge University Press.

Gentzkow, M. and Shapiro, J. M. (2010). What drives media slant? evidence from us daily newspapers. *Econometrica*, 78(1):35–71.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.

Gentzkow, M., Shapiro, J. M., and Taddy, M. (2017). Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech.

Hansen, C. and Kozbur, D. (2014). Instrumental variables estimation with many weak instruments using regularized jive. *Journal of Econometrics*, 182(2):290–308.

Hartford, J., Lewis, G., Leyton-Brown, K., and Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, pages 1414–1423.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Larning*. Springer, New York, NY.

Hopkins, D. J., Ladd, J. M., et al. (2014). The consequences of broader media choice: Evidence from the expansion of fox news. *Quarterly Journal of Political Science*, 9(1):115–135.

Huang, A. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*, volume 4, pages 9–56.

Kusner, M., Sun, Y., Kolkin, N., and Weinberger, K. (2015). From word embeddings to document distances. In *International Conference on Machine Learning*, pages 957–966.

Lee, M. D., Pincombe, B., and Welsh, M. (2005). An empirical evaluation of models of text document similarity. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 27.

Li, M., Chen, X., Li, X., Ma, B., and Vitányi, P. (2003). The similarity metric. In *Proceedings* *of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 863–872. Society for Industrial and Applied Mathematics.

Lucas, C., Nielsen, R. A., Roberts, M. E., Stewart, B. M., Storer, A., and Tingley, D. (2015). Computer-assisted text analysis for comparative politics. *Political Analysis*, 23(2):254–277.

Martin, G. and Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*.

Nakov, P., Popova, A., and Mateev, P. (2001). Weight functions impact on lsa performance. *EuroConference RANLP*, pages 187–193.

Pincombe, B. (2004). Comparison of human and latent semantic analysis (lsa) judgements of pairwise document similarities for a news corpus. Technical report, DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION SALISBURY (AUSTRALIA) INFO âŠ.

Poole, K. T. and Rosenthal, H. (2001). D-nominate after 10 years: A comparative update to congress: A political-economic history of roll-call voting. *Legislative Studies Quarterly*, pages 5–29.

Prat, A. (2017). Media power. *Journal of Political Economy*.

Schroeder, E. and Stone, D. F. (2015). Fox news and political knowledge. *Journal of Public Economics*, 126:52–63.

Snyder, J. and Stromberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2):355–408.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529.

A Robustness analysis: Similarity score

A.1 Weights

Our original formula of the cosine similarity score (1), for the sake of parsimony, did not include any weights. Meanwhile, the weights are a crucial part of the general version of this formula:

$$\text{sim}(obs_a = A, obs_b = B) = cos(AB) = \frac{\sum_{i=1}^{n} w_{Ai} A_i w_{Bi} B_i}{\sqrt{\sum_{i=1}^{n} (w_{Ai} A)_i^2} \sqrt{\sum_{i=1}^{n} (w_{Bi} B)_i^2}} \tag{5}$$

Weighting the features based on their statistical properties is, probably, the most transparent way to improve the cosine similarity measure without adding any exogenous assumptions. The purpose of this transformation is to enhance the underlying semantic structure of the quantified texts. For our task, we construct the weights independently for each of our corpora, three cable channels, and U.S House transcripts, to emphasize the term importance within each corpus.

The weighting function consists of the local component, L(j,i), that captures the importance of a feature within a single document and the global one, G(i), to account for the importance within the whole corpus:

$$w_{ji} = L(j,i)G(i) \tag{6}$$

Based on the existing literature on the topic (Pincombe (2004), Nakov et al. (2001), Lee et al. (2005), etc)), we pick the most widely used and discussed weighting functions, three local and four global. Then, we investigate how the main result of our analysis – the effect of the Fox rating on the similarity score - changes dependent on the weights.

Table A.1: Local weighting functions

	(1)	(2)	(3)
L(j,i)	1	$tf(j,i) = \frac{f(j,i)}{\sum f(j,k)}$	$log(1 + tf(j,i))$

Table A.2: Global weighting functions

	(1)	(2)	(3)	(4)
$G(i)$	1	$idf(d\|i) = log\frac{\|i\|}{\|d \in D: i \in d\|}$	$H(d\|i) = -\sum p(i,j)log(p(i,j))$	$1 - H(d\|i)/H(d)$

For the local weight functions (LWF), we try the constant (1), the term frequency or tf (2), and the logarithmic term-frequency or log-tf (3). For the global one, we look at the constant (1), the inverse Document frequency or idf (2), the entropy (3), and the real entropy of the conditional distribution (4), where $p(j,i) = \frac{tf(j,i)}{gf(i)}$ and $gf(i)$ is the global frequency of the term.

Tables (A.3) and (A.4) show the results for the Fox news' effect on the similarity score dependent on various weights. The first columns in both tables show the results with no weights as a benchmark. They were already presented in the previous subsection. We see that the models' estimates are robust: the magnitude and the Kleibergen-Paap F do not experience significant fluctuation with the exception for the second model. The tf-idf model has the highest Kleibergen-Paap F-statistic. However, the effect's magnitude's significance lowers to the 10% level. Based on both tables, it looks that the log-tf - real entropy model provides the highest estimate for the effect of interest.

Table A.3: 2SLS Estimates with various feature weights: Fox (no demographics)

$L(j,i)G(i)$	(1)(1)	(2)(2)	(2)(3)	(3)(3)	(2)(4)	(3)(4)
	2SLS Effect on Sim to Fox					
Fox Ratings	0.363**	0.330*	0.363**	0.364**	0.368**	0.368**
	(0.183)	(0.178)	(0.183)	(0.183)	(0.186)	(0.186)
Observations	1,398	1,396	1,398	1,398	1,397	1,397
Kleibergen-Paap F	11.75	11.79	11.75	11.75	11.76	11.76

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

These results provide additional proof to the discovered effect of the Fox rating: its estimate is strong enough not to be sensitive to a particular weighting scheme. Also, the results defend the choice of our approach to constructing the features. They all have similar statistical importance, since various weights almost do not affect the cosine similarity's estimates.

Similar tables for CNN and MSNBC are available upon request.

Table A.4: 2SLS Estimates with various feature weights: Fox (demographic controls)

$L(j,i)G(i)$	(1)(1)	(2)(2)	(2)(3)	(3)(3)	(2)(4)	(3)(4)
	2SLS Effect on Sim to Fox					
Fox Ratings	0.454**	0.420*	0.454**	0.456**	0.472**	0.473**
	(0.228)	(0.219)	(0.228)	(0.228)	(0.232)	(0.232)
Observations	1,398	1,396	1,398	1,398	1,397	1,397
Kleibergen-Paap F	8.827	8.889	8.827	8.827	8.836	8.836

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

A.2 Dropping features

Another possible robustness test is based on randomly dropping the features, 3-grams in our case, from our corpora. Since, our strategy of selecting features is already pretty conservative, we do not expect the estimate to stay significant when the number of features gets smaller. In the meanwhile, in the case of the robustness of our measure, we do expect the sign of the effect to stay the same.

Before looking at the results, one important thing to keep in mind is that the results of this test are a random trial. When we randomly drop certain features the effect's estimate may even increase. Table (A.5) shows this anomaly. However, most importantly, we see the robustness of the sign of the effect while we keep more than 50% of our features.

Table A.5: 2SLS Estimates with features dropped: Fox (demographic controls)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
share	100%	99%	90%	80%	70%	50%	30%	10%
Fox Rating	0.454**	0.364*	0.376*	0.686*	0.223	0.287	-0.724	-0.0704
	(0.228)	(0.220)	(0.210)	(0.405)	(0.308)	(0.279)	(1.111)	(0.607)
Observations	1,398	1,396	1,388	1,338	1,288	1,226	378	186
Kleibergen-Paap F	8.827	8.721	8.596	7.644	6.168	6.998	0.800	3.035

Robust standard errors in parentheses, clustered by district
*** p<0.01, ** p<0.05, * p<0.1

B Heterogeneous Effects Across Congressmen

Here we ask whether there are heterogeneous effects of Fox News Channel. First, we are interested in whether there are different effects across parties. The previous papers looking at how Fox News affects elections find that voter effects are driven by higher turnout by Republicans, rather than persuading Democrats (DellaVigna and Kaplan, 2007). We show in Table B.6 that there are not large differences in effects across political parties (Columns 1 and 2).

Second, we are interested in whether there are differences in the effect by the seniority of the Congressman. We find that the effect is concentrated among less-experienced Congressmen (Column 3). This could mean that these congressmen have a weaker incumbency

Table B.6: Heterogeneous Effects of Fox News Channel

	(1)	(2)	(3)	(4)
	2SLS Effect on Sim to Fox			
Fox Ratings	0.325*	0.412	0.871*	0.135
	(0.178)	(0.541)	(0.475)	(0.253)
Sample	Dems	Repubs	Young	Old
Observations	676	688	718	634
Kleibergen-Paap F	10.25	2.526	4.235	3.867

Robust standard errors in parentheses, clustered by district.
*** p<0.01, ** p<0.05, * p<0.1

Table C.7: Summary Statistics on Demographics (2000 Census)

Variable	N	Mean	SD	Min	Max
Log Population	1398	9.88	0.3	8.68	10.88
Income Below 25K (%)	1398	27.96	8.19	8.63	57.07
Income 25K-49K (%)	1398	29.18	3.77	17.49	35.42
Income 50K-99K (%)	1398	30.19	4.99	13.65	43.41
Income 100K-199K (%)	1398	10.2	5.27	2.18	27.9
Income 200K+ (%)	1398	2.46	1.98	0.66	15.49
Hispanic Share (%)	1398	18.94	26.27	0.82	167.91
Black Share (%)	1398	18.61	20.8	0.39	114.16
Age 70+ (%)	1398	9.37	2.64	3.58	21.43
Age 55-69 (%)	1398	12.19	1.71	7.45	21.7
Age 35-54 (%)	1398	29.58	2.08	23.3	35.94
Age 15-34 (%)	1398	27.58	2.92	17.6	39.71
Age 0-14 (%)	1398	21.28	2.26	12.7	29.33
Grad School (%)	1398	8.7	3.65	2.29	25.58
Bachelor (%)	1398	15.48	5.1	4.65	33.29
Associates (%)	1398	6.31	1.44	2.46	10.23
Some College (%) 0	1398	21.03	3.47	11.3	29.52
High School (%)	1398	28.95	6.31	13.51	46.25
Some High School (%)	1398	12.05	3.6	4.47	25.84
Less than 9th grade (%)	1398	7.49	4.56	1.86	39.52
Renter (%)	1398	30.08	10.51	13.41	85.01
Homeowner (%)	1398	62.15	9.91	9.19	80.25
Log Median HH Income (%)	1398	10.7	0.24	9.95	11.41
Log Per Capita Income (%)	1398	2.3	0.02	2.23	2.38
Log Population Density	1398	7.25	1.32	3.87	11.32

advantage, so the political attitudes of their constituents (shifted by Fox) have a bigger impact. These politicians might also have a less entrenched policy agenda and therefore are more easily persuaded.

C Additional Tables and Figures

Table C.8: Trigrams Most Associated with each Cable News Network

n3gram	FOX Assoc.	n3gram	CNN Assoc.	n3gram	MSBNC Assoc.
bush_tax_cut	1.497	payrol_tax_cut	3.000	presid_georg_bush	1.334
suprem_court_rule	1.346	rais_debt_ceil	1.888	presid_unit_state	1.288
state_suprem_court	1.306	health_care_law	1.809	health_care_plan	1.266
suprem_court_justic	1.216	town_hall_meet	1.648	unit_state_senat	1.252
health_care_law	1.191	depart_homeland_secur	1.605	weapon_mass_destruct	1.247
doesnt_make_sens	1.166	health_care_cost	1.502	world_trade_center	1.244
american_peopl_want	1.141	health_care_system	1.500	wall_street_journal	1.226
senat_major_leader	1.128	congression_budget_offic	1.453	state_union_address	1.147
senat_judiciari_committe	1.123	osama_bin_laden	1.354	health_care_bill	1.127
everi_singl_day	1.097	stem_cell_research	1.304	foreign_relat_committe	1.126

Figure C.1: Scree Plot for Principal Components of Demographic Characteristics

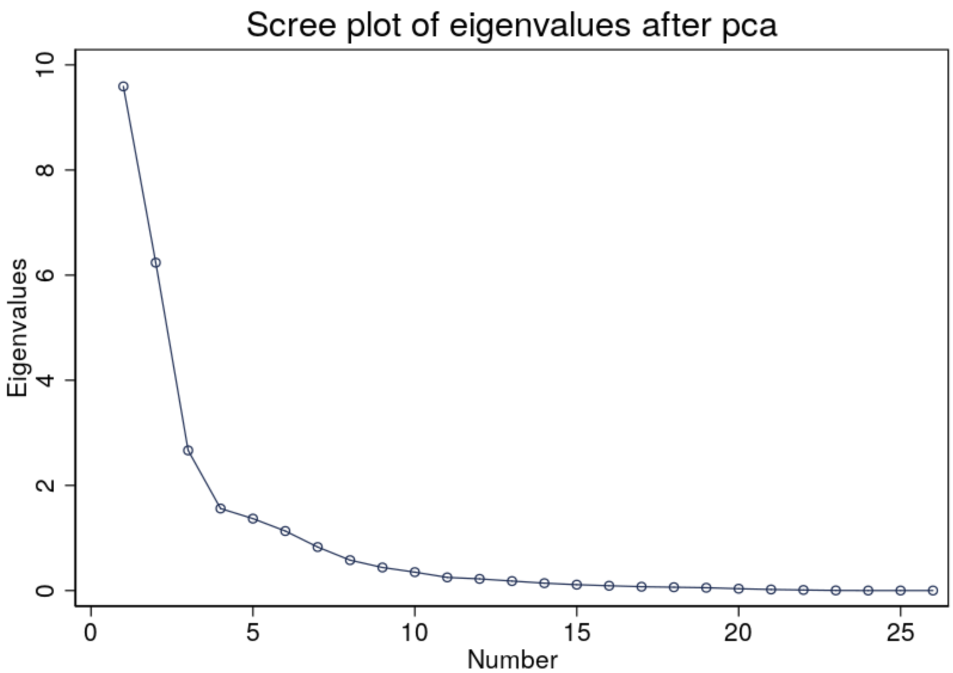


Figure C.2: True Reduced Form vs. 1995 Placebo Reduced Form