Online Appendix
for
Does the Media Cover the Economy Accurately? An Analysis of Sixteen Developed Democracies*

Mark A. Kayser†
Michael Peress‡

Forthcoming in the *Quarterly Journal of Political Science*

Abstract

Can voters learn what they need to learn to hold governments accountable for the economy through news coverage? Employing the first large-scale cross-national dataset of media coverage of the economy—over 2 million articles related to three economic indicators in 32 mainstream newspapers, one left-wing and one right-wing, in 16 developed countries and six languages—we investigate media coverage of the economy that bears implications for electoral accountability and partisan advantage. We find that the tone of most mainstream newspapers tracks the economy faithfully, although the frequency of coverage increases with negative outcomes. While we find some evidence for partisan bias in tone for growth headlines and in frequency of coverage for unemployment articles, its substantive magnitude is diminutive. Mainstream newspaper coverage of the economy provides voters with largely accurate information.

*We gratefully acknowledge peerless research assistance from Giuliana Pardelli, CJ Yetman, Jeff Arnold, Arndt Leininger, William Pollock, and Johanna Willmann who all worked on data collection. Special thanks go to Giuliana Pardelli for her dictionary work in five languages, Normand Peladeau for repeated dictionary guidance, and to CJ Yetman for innovative automation. Joia Buning, Mujahed Islam, Sven Rudolf, Loïc Baptiste Savatier, Viviane Schilling, Bruno-Pierre St-Jacques and Bas Vervaart all assisted with for human coding for validity checks. We thank John Ahlquist, Mike Annany, Ryan Enos, Paul Kellstedt, and Marc Meredith for helpful feedback as well as participants of seminars at the Center for Advanced Studies in the Behavioral Sciences at Stanford, Harvard, NYU, NYU Abu Dhabi, UC Berkeley, the University of Bremen, various meetings of the American, European and Midwest Political Science Associations, and the Workshop in Political Economy and Political Science (Santiago).

†Hertie School of Governance, Berlin, kayser@hertie-school.org
‡SUNY-Stony Brook, michael.peress@stonybrook.edu

A Online Appendix

A.1 Human Coding

In this appendix, we provide further details on the human coding. We applied human coding to three languages covering 13 of the 16 countries in our sample. Two research assistant coded approximately 1500 English articles and headlines, 750 German articles and headlines, and 500 French articles and headlines. For both articles and headlines, the coders coded items on the overall economy, growth, unemployment, and inflation. In each case, the coders coded the article on a 5-point scale (strongly negative, weakly negative, neutral, weakly positive, and strongly positive), or indicated that the article or headline was not substantially about the economy, growth, unemployment, and inflation, respectively.

Table A1 report the reliability of the coding. For the topic of the articles, the error rate was quite small in most cases. Identifying growth articles consistently proved to be the hardest task. For coding sentiment, the coders did not always select the same point on the 5 point scale, but it was rare to find a pure error—where one coder coded the article/headline as positive and the other coded it as negative.

A.2 Checking the Measures

Each monthly measure of sentiment is based on a fraction of economic words that are near positive rather than negative words. The standard error for this proportion is given by $SE_j = \sqrt{\frac{p_j(1-p_j)}{W_j}}$ where p_j is the proportion of economic words near positive words and W_j is the number of economic words in newspaper-month j. To get a sense of how noisy our measure of sentiment is, we report the average, the 2.5%, and the 97.5% quantiles for each of the 32 newspapers in our sample. These results are reported in Table A2. The amount of measurement error varies quite a bit, with the German and Luxembourgian papers having the most measurement error and the U.S. papers having the least measurement error.

	All		English		French		German	
	Articles	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*
Economic Articles								
Topic:								
Both Yes	0.450	0.358	0.499	0.437	0.548	0.412	0.290	0.161
Both No	0.428	0.521	0.359	0.419	0.336	0.508	0.624	0.738
Error	0.122	0.121	0.142	0.144	0.116	0.080	0.086	0.101
Kappa	0.757	0.754	0.709	0.712	0.765	0.836	0.808	0.723
n	2696	2770	1448	1528	500	498	748	744
Sentiment:								
Same	0.775	0.795	0.779	0.786	0.772	0.839	0.763	0.769
Same Direction	0.062	0.074	0.061	0.074	0.067	0.070	0.059	0.083
One Neutral	0.110	0.100	0.092	0.107	0.147	0.059	0.129	0.130
Error	0.053	0.031	0.067	0.033	0.013	0.032	0.048	0.019
Kappa	0.666	0.725	0.670	0.707	0.663	0.797	0.645	0.701
n	1062	902	652	608	224	186	186	108
Growth Articles								
	All		**English**		**French**		**German**	
	Articles	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*
Topic:								
Both Yes	0.400	0.330	0.416	0.339	0.435	0.339	0.247	0.147
Both No	0.404	0.487	0.411	0.484	0.361	0.521	0.407	0.441
Error	0.197	0.183	0.173	0.177	0.204	0.141	0.345	0.412
Kappa	0.606	0.628	0.654	0.641	0.591	0.672	0.202	0.175
n	1714	1548	1290	1288	230	192	194	68
Sentiment:								
Same	0.765	0.784	0.762	0.772	0.778	0.875	0.773	0.900
Same Direction	0.082	0.093	0.074	0.090	0.111	0.125	0.136	0.100
One Neutral	0.079	0.086	0.081	0.097	0.089	0.000	0.000	0.000
Error	0.074	0.036	0.083	0.041	0.022	0.000	0.091	0.000
Kappa	0.693	0.749	0.668	0.711	0.776	1.000	0.812	1.000
n	582	440	470	390	90	40	22	10
Unemployment Articles								
	All		**English**		**French**		**German**	
	Articles	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*
Topic:								
Both Yes	0.134	0.050	0.088	0.028	0.284	0.139	0.474	0.600
Both No	0.775	0.931	0.828	0.957	0.608	0.835	0.382	0.200
Error	0.091	0.019	0.085	0.015	0.108	0.026	0.145	0.200
Kappa	0.689	0.823	0.667	0.777	0.737	0.875	0.579	0.482
n	1576	1494	1278	1280	222	194	76	20
Sentiment:								
Same	0.793	0.781	0.795	0.867	0.778	0.682	0.812	0.750
Same Direction	0.069	0.109	0.057	0.100	0.074	0.091	0.094	0.167
One Neutral	0.075	0.078	0.057	0.033	0.130	0.182	0.031	0.000
Error	0.063	0.031	0.091	0.000	0.019	0.045	0.062	0.083
Kappa	0.692	0.773	0.660	0.932	0.625	0.450	0.812	0.833
n	174	64	88	30	54	22	32	12
Inflation Articles								
	All		**English**		**French**		**German**	
	Articles	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*	*Articles*	*Headlines*
Topic:								
Both Yes	0.056	0.018	0.057	0.017	0.027	0.026	0.175	0.000
Both No	0.890	0.963	0.882	0.962	0.959	0.969	0.750	1.000
Error	0.054	0.019	0.061	0.021	0.014	0.005	0.075	0.000
Kappa	0.637	0.612	0.639	0.567	0.799	0.895	0.387	1.000
n	1544	1462	1282	1266	222	194	40	2
Sentiment:								
Same	0.794	0.714	0.804	0.700	0.667	0.750	0.833	
Same Direction	0.088	0.000	0.107	0.000	0.000	0.000	0.000	
One Neutral	0.088	0.214	0.071	0.300	0.167	0.000	0.167	
Error	0.029	0.071	0.018	0.000	0.167	0.250	0.000	
Kappa	0.749	0.293	0.812	0.062	0.183	0.500	-0.174	
n	68	14	56	10	6	4	6	0

Table A1: Coder Reliability. "Both Yes" and "Both No" indicate that both coders made the same choice in assigning an article to a topic. "Same" indicates that coders reported the same value on the five point scale. "Same Direction" indicates that both coders selected a positive rating, both coders selected a negative rating, or both coders selected a neutral rating. "One Neutral" indicates that one coder select a neutral rating with the other selecting a positive or negative rating. "Error" indicates that one coder assigned an article to a particular topic while the other did not, or that one coder selected a positive rating while the other selected a negative rating. *Kappa* is Cohen's (1960) measure of inter-coder reliability.

	Mean S.D.	2.5% Quantile of S.D.	97.5% Quantile of S.D.
The Age (Australia)	0.011	0.008	0.015
Herald Sun (Australia)	0.018	0.011	0.021
Der Standard (Austria)	0.056	0.026	0.086
Die Presse (Austria)	0.051	0.028	0.079
Toronto Star (Canada)	0.012	0.007	0.016
The Globe and Mail (Canada)	0.011	0.008	0.013
Le Monde (France)	0.036	0.024	0.060
Le Figaro (France)	0.037	0.024	0.066
Die Zeit (Germany)	0.067	0.031	0.105
Frankfurter (Germany)	0.084	0.048	0.141
Irish Times (Ireland)	0.012	0.009	0.018
Irish Independent (Ireland)	0.016	0.010	0.022
Globes (Israel)	0.033	0.022	0.050
Jerumselum Post (Israel)	0.022	0.015	0.032
La Stampa (Italy)	0.020	0.016	0.032
Cor (Italy)	0.016	0.012	0.022
Nikkei (Japan)	0.024	0.017	0.034
Yomiuri (Japan)	0.026	0.017	0.038
Le Quotidien (Luxembourg)	0.071	0.035	0.126
Age Fi (Luxembourg)	0.058	0.036	0.154
New Zealand Press	0.028	0.017	0.074
New Zealand Herald	0.018	0.012	0.027
Cor (Portugal)	0.052	0.033	0.091
Noticias (Portugal)	0.048	0.028	0.088
El Pais (Spain)	0.024	0.015	0.037
El Mundo (Spain)	0.031	0.015	0.051
Tages (Switzerland)	0.072	0.033	0.111
NZZ (Switzerland)	0.036	0.017	0.048
Guardian (U.K.)	0.015	0.010	0.021
London Times (U.K.)	0.014	0.008	0.019
New York Times (U.S.)	0.009	0.007	0.011
Wall Street Journal (U.S.)	0.011	0.006	0.021

Table A2: Sampling Error in Dictionary Coding

The above table implicitly assumed two things—that each pairing of a positive/negative word with an economic word in the text produces an unbiased estimate of sentiment and that the errors were independent across such sentence fragments. Since these assumptions may be violated, the table can be interpreted as a lower bound on the level of measurement error in our monthly measure of sentiment. To get a better idea of the amount of measurement error, we relied on human coding of a sample of articles as a benchmark. The most straightforward comparison would be to code all articles from a random sample of months and use this to construct a monthly measure of human-coded sentiment, but this is not feasible—even coding a sample of 100 months would require coding about 50,000 articles. Instead, we human-coded a random sample of articles and headlines and extrapolated from this sample to measure the error in our monthly estimates.

We had a team of trained research assistants code the articles and (separately) headlines in three of the languages according to the following scheme. Each coder was instructed to code the coverage and tone of the overall economy, growth, unemployment, and inflation for a series of articles and headlines. Each item was coded based on whether the article/headline was strongly negative, weakly negative, neutral, weakly positive, strongly positive, or not applicable. For example, an article would be coded as not applicable on inflation if it was not substantially about inflation. The articles and headlines coded were a stratified random sample. The sample was constructed so that country-days were sampled first and within each country-day, two articles were sampled from the left-wing and right-wing newspapers from that country. Each article was then coded by two coders. The sampling scheme was designed to allow for (i) a measure of reliability between two coders, (ii) a measure of the correlation of coding errors within days, and (iii) for a direct analysis of coverage and bias on the hand-coded articles (considered later as a robustness check on our main results).

For comparison, we were able to compute our dictionary-coded measure at the article level. At the article-level, our measure is very error prone because it is based on a small

number of sentence fragments (typically between 1 and 2) and because dictionary-coded measures cannot perfectly capture the meaning of human language. For example, the dictionary measure will code "The economy grew by 5%" as positive sentiment on growth based on the proximity of "grew" to "economy". Similarly, it will code "December was not a bad month for unemployment" as positive unemployment because of the proximity of "bad" to "unemployment", with the meaning of "bad" modified by its proximity to "not". In both cases, the dictionary-coded measures correspond to the meaning a human reader would likely extract from the text fragment. At the same time, one can cook up examples that our dictionary will code in ways that depart for the meaning a human reader would extract. "The economy only grew by 0.1%" would be coded as positive growth sentiment. Similarly, "South Korea's economy grew by over 10% every year in the 1970s" would be coded as positive growth sentiment for a given month in 2010 if written in the *Wall Street Journal* in that month in 2010. For dictionary measures to preform well, the language used in newspapers must be such that these types of errors are rare, or average out, when aggregated over a month. There is no way to know whether this is the case without empirically testing it.

As a preliminary test, we ran a series of logit and ordered logit models where human-coded coverage of sentiment was the dependent variable and the article-level dictionary measure of coverage of sentiment was the independent variable. These results are given in Table A3. The results indicate a positive relationship between the dictionary-coded measures and the human coding of coverage and sentiment.

We can use these results to obtain an estimate of the amount of measurement error that accounts for the fact that the dictionary-coded measure may not be unbiased on average and the errors may not be independent. We accomplished this using a random effects model. Let s_l denote the true sentiment for article l and let $\tilde{s}_{li}$ be the human-coded sentiment for article l by coder i. We modeled $\tilde{s}_{li}$ using an ordered probit specification. There is a latent variable $\tilde{s}^*_{li} = s_l + \varepsilon^R_{li}$ where $\varepsilon^R_{li} \sim N(0, \sigma_R)$ are i.i.d. We assume that we observe

				Topics				
		Articles				**Headlines**		
	Economy	*Growth*	*Unem.*	*Inflation*	*Economy*	*Growth*	*Unem.*	*Inflation*
Slope	93.34***	82.73***	314.07***	132.12***	11.65***	4.81**	35.26***	22.38***
	(8.63)	(12.26)	(38.50)	(16.49)	(1.58)	(1.82)	(7.75)	(2.55)
N	2966	2124	2026	2007	3080	1893	1838	1818
				Sentiment (Logit model)				
		Articles				**Headlines**		
	Economy	*Growth*	*Unem.*	*Inflation*	*Economy*	*Growth*	*Unem.*	*Inflation*
Slope	0.63**	0.44+	2.05***	1.30*	2.85***	3.01**	3.51**	42.72***
	(0.22)	(0.23)	(0.50)	(0.65)	(0.62)	(1.06)	(1.29)	(2.93)
N	851	507	172	107	103	47	24	9
				Sentiment (Ordered Logit model)				
		Articles				**Headlines**		
	Economy	*Growth*	*Unem.*	*Inflation*	*Economy*	*Growth*	*Unem.*	*Inflation*
Slope	0.54**	0.35+	1.72***	0.93	2.50***	2.54***	3.23**	6.75
	(0.18)	(0.20)	(0.40)	(0.57)	(0.51)	(0.76)	(1.05)	(4.31)
N	978	558	184	121	119	52	25	11

Table A3: Comparing Human Coding to Dictionary Coding. In each case, the dependent variable is the human-coded measure and the independent variable is the dictionary-based measure. Constant terms and cutpoints are omitted from the table. Standard errors in parentheses. $^{+}p < .10,^{*}p < .05,^{**}p < .01,^{***}p < .001$.

$\tilde{s}_{li} \in \{1, 2, 3, 4, 5\}$ depending on where the latent variable falls relative to 4 cutpoints. We modeled the dictionary-coded measure as $s_l = a + b s_l + \varepsilon_l^L + \varepsilon_d^D(l) + \varepsilon_m^M(l)$ where $\varepsilon_l^L \sim N(0, \sigma_L)$, $\varepsilon_l^D \sim N(0, \sigma_D)$, $\varepsilon_l^M \sim N(0, \sigma_M)$, $d(l)$ indicates the day of article l, and $m(l)$ denotes the month of article l. The model allows for the errors of the dictionary-coded measures be correlated within days and within months. It also allows for the possibility that the dictionary-coded measure may be poor if b is small relative to the size of the errors.

We estimated the model above using Simulated Maximum Likelihood. Our estimates indicate that $\hat{\sigma}_L = 0.323$ with a standard error of (0.013), $\hat{\sigma}_D = 0.032$ with a standard error of (0.111), and $\hat{\sigma}_M = 0.006$ with a standard error of (0.015). We find that the idiosyncratic error is the largest component of the error, but there is a degree of error that is correlated within the same day. This suggests that the error of the dictionary-coded measure will not completely disappear if we have an infinite number of articles within each month. Under the assumption that the number of articles in each day of the same month and that there are 30 days in each month, we can calculate that the measurement error will be $SE_j = \sqrt{\sigma_L^2/L_j + \sigma_D^2/30 + \sigma_M^2}$

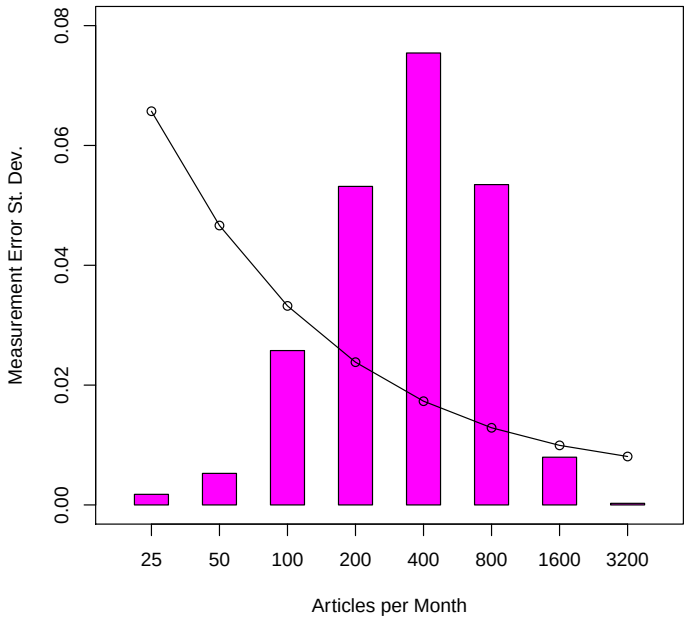

Figure A1: Measurement Error in Sentiment. The line represents the amount of measurement error in the monthly estimates of sentiment implied by the model. The magenta bars represent the distribution of articles per month. The scale on the left is for the line representing the amount of measurement error.

where L_j is the number of articles in newspaper-month j. Figure A1 presents these results along with the distribution of articles per month in our dataset. For most months, the standard error of our measure of sentiment will range between 0.015 and 0.035. In months where there is a very large number of articles, the standard deviation can go to as low as 0.008. Since the scale ranges between 0 and 1, this suggests that our measure is relatively accurate. Beyond this, the results suggest that while dictionary coding does depart from actual meaning on a sentence fragment level, averaging over a month leads to relatively precise estimates of sentiment.

A.3 Additional Figures and Tables

Dependent Variable:	Economic Sentiment (headlines)	Growth Sentiment (headlines)	Unemployment Sentiment (headlines)	Inflation Sentiment (headlines)
Independent Variables:				
Constant	0.390***	0.370***	0.437***	0.425***
	(0.013)	(0.019)	(0.014)	(0.007)
Growth (yearly)	0.018***	0.029***	0.016***	-0.002
(SD = 3.052)	(0.003)	(0.004)	(0.003)	(0.002)
Change in Unem. (yearly)	0.000	-0.017	-0.010+	0.008+
(SD = 0.930)	(0.004)	(0.011)	(0.006)	(0.005)
Change in Inf. (yearly)	-0.004**	-0.006**	0.003+	-0.011***
(SD = 4.062)	(0.001)	(0.002)	(0.002)	(0.003)
Effect Size:	0.616***	0.750***	0.582***	0.505***
	(0.028)	(0.048)	(0.039)	(0.026)
Number of Months	6402	6087	5345	5672
Number of Newspapers	31	31	31	31
Number of Countries	16	16	16	16
R-Squared	0.060	0.101	0.031	0.016

Table A4: The Effect of the Economy on Newspaper Sentiment in Headlines. Newspaper fixed effects were included in each regression, but omitted from the table. Standard errors clustered by newspaper in parentheses. $^{+}p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

Dependent Variable:	Economic Coverage (articles)	Growth Share of Coverage (articles)	Unemployment Share of Coverage (articles)	Inflation Share of Coverage (articles)
Independent Variables:				
Growth (yearly)	-0.001***	-0.017***		
	(0.000)	(0.002)		
Unemployment (yearly)	0.000		0.015***	
	(0.000)		(0.003)	
Inflation (yearly)	0.000			0.010***
	(0.000)			(0.003)
Number of Months	6136	6402	6407	6409
Number of Newspapers	32	31	31	31
Number of Countries	16	16	16	16
R-Squared	0.739	0.306	0.230	0.309

Table A5: The Effect of the Economy on Newspaper Coverage. Newspaper fixed effects were included in each equation, but omitted from the table. Standard errors clustered by newspaper in parentheses. $^{+}p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

Dependent Variable:	Growth Sentiment (cong. rec.)	Unemployment Sentiment (cong. rec.)	Inflation Sentiment (cong. rec.)	Growth Coverage (cong. rec.)	Unemployment Coverage (cong. rec.)	Inflation Coverage (cong. rec.)
Independent Variables:						
Growth (yearly)	0.017***	-0.002	0.009**	-0.009**		
	(0.002)	(0.003)	(0.003)	(0.004)		
Unemployment (yearly)					0.027***	
					(0.003)	
Change in Unemployment (yearly)		-0.008				
		(0.006)				
Inflation (yearly)						0.012
						(0.007)
Change in Inflation (yearly)			-0.007***			
			(0.002)			
Ideological Match	0.032**	0.055***	0.029*	-0.001	-0.102***	-0.033
	(0.010)	(0.013)	(0.014)	(0.017)	(0.027)	(0.029)
Growth * Ideo. Match	-0.006*	-0.007+	-0.006	0.005		
	(0.003)	(0.004)	(0.004)	(0.005)		
Unem. * Ideo. Match					0.019***	
					(0.005)	
Change in Unem. * Ideo. Match		-0.003				
		(0.007)				
Inf. * Ideo. Match						0.004
						(0.010)
Change in Inf. * Ideo. Match			0.006+			
			(0.003)			
p-Value from Wald Test:						
Ideology Terms = 0	0.005**	0.000***	0.051+	0.156	0.000***	0.265
Number of Months	385	325	385	416	420	422
R-Squared	0.143	0.099	0.059	0.024	0.384	0.024

Table A6: Media Bias in Economic Sentiment in the Congressional Record. Robust standard errors in parentheses. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interaction with the economy are jointly zero. $^{+}p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

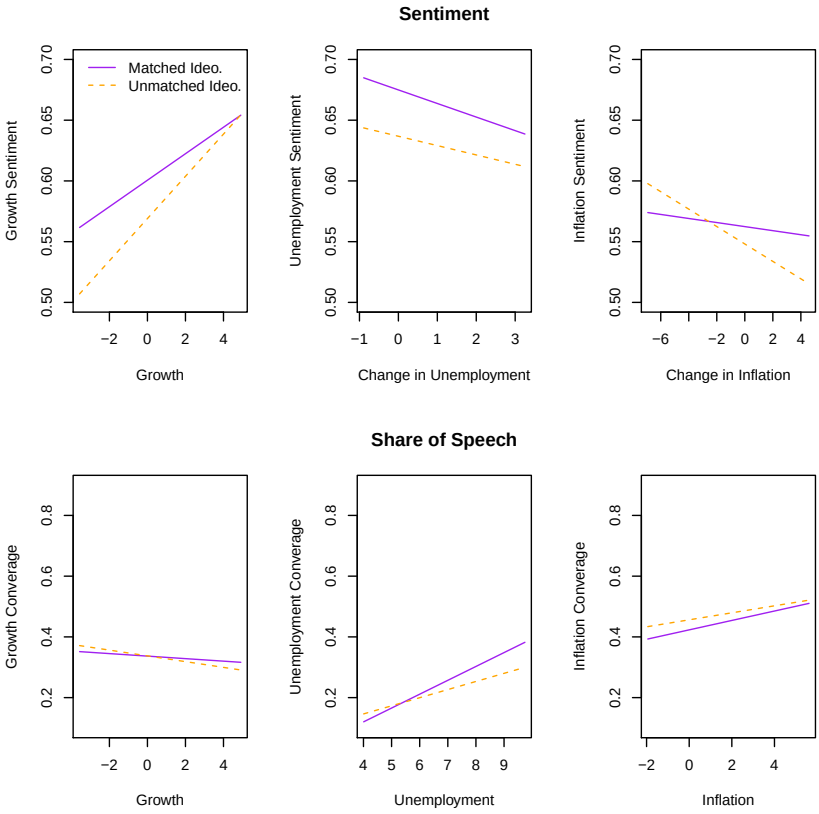

Figure A2: Sentiment and Coverage vs. the Economy for Ideologically Matched and Unmatched U.S. Senators. Results are calculated based on Table A6

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)
Independent Variables:						
Growth (yearly)	0.030***	0.009*	-0.005*	-0.011***		
	(0.005)	(0.004)	(0.002)	(0.003)		
Unemployment (yearly)					0.014***	
					(0.004)	
Change in Unemployment (yearly)		-0.013				
		(0.009)				
Inflation (yearly)						0.011***
						(0.003)
Change in Inflation (yearly)			-0.014**			
			(0.004)			
Ideological Match	0.033	0.003	-0.003	0.003	0.003	0.015
	(0.026)	(0.015)	(0.012)	(0.015)	(0.037)	(0.023)
Growth * Ideo. Match	-0.005	-0.002	0.002	-0.004		
	(0.007)	(0.005)	(0.003)	(0.003)		
Unem. * Ideo. Match					-0.001	
					(0.005)	
Change in Unem. * Ideo. Match		-0.007				
		(0.013)				
Inf. * Ideo. Match						0.000
						(0.006)
Change in Inf. * Ideo. Match			0.009*			
			(0.005)			
p-Value from Wald Test:						
Ideology Terms = 0	0.410	0.917	0.153	0.441	0.871	0.723
Number of Months	988	986	979	988	990	990
Number of Newspapers	31	31	31	31	31	31
Number of Countries	16	16	16	16	16	16
R-Squared	0.507	0.282	0.353	0.653	0.636	0.702

Table A7: Bias in Economic Sentiment and Coverage within Six Months of an Election. Newspaper fixed effects were included in the analysis, but omitted from the table. Standard errors clustered by newspaper are in parentheses. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interaction with the economy are jointly zero. $^{+}p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

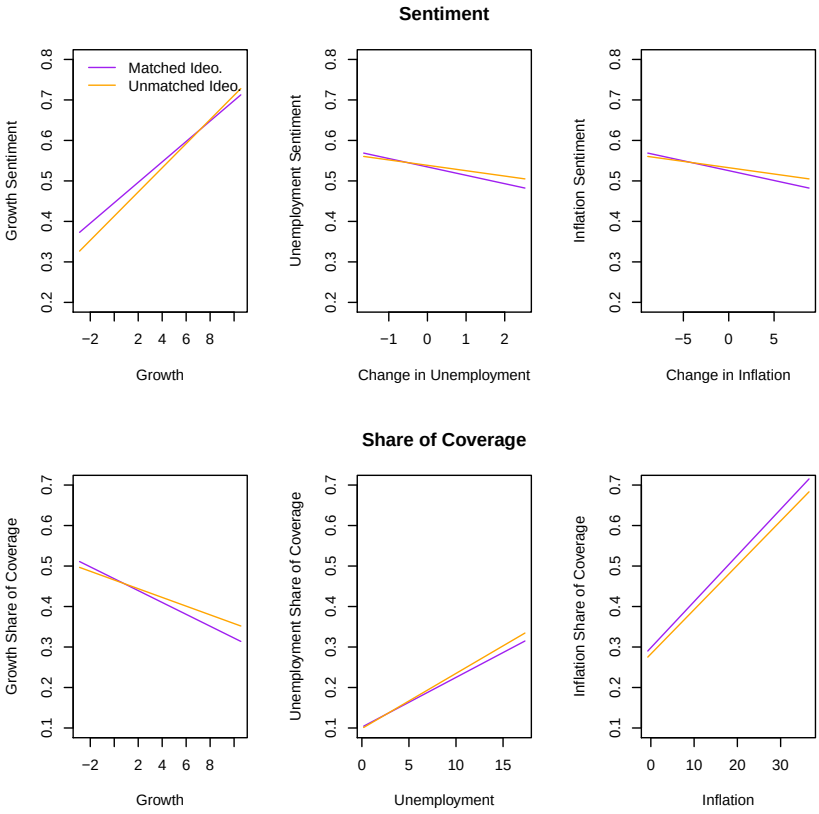

Figure A3: Sentiment and Coverage vs. the Economy for Ideologically Matched and Unmatched Newspapers during Election Campaigns. Results are calculated based on Table A7

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)
Independent Variables:						
Growth (yearly)	0.031***	0.012***	-0.005*	-0.018***		
	(0.006)	(0.003)	(0.002)	(0.002)		
Unemployment (yearly)					0.014***	
					(0.002)	
Change in Unemployment (yearly)		-0.015**				
		(0.005)				
Inflation (yearly)						0.010***
						(0.003)
Change in Inflation (yearly)			-0.011***			
			(0.002)			
Ideological Match	-0.001	0.001	0.001	0.003	0.027	0.001
	(0.022)	(0.011)	(0.009)	(0.014)	(0.018)	(0.018)
Growth * Ideo. Match	0.003	0.002	-0.001	-0.003		
	(0.007)	(0.003)	(0.003)	(0.004)		
Unem. * Ideo. Match					-0.003	
					(0.002)	
Change in Unem. * Ideo. Match		-0.006				
		(0.007)				
Inf. * Ideo. Match						-0.001
						(0.005)
Change in Inf. * Ideo. Match			0.001			
			(0.003)			
p-Value from Wald Test:						
Ideology Terms = 0	0.783	0.683	0.928	0.560	0.330	0.982
Number of Months	6094	6026	5870	6098	6106	6109
Number of Newspapers	31	31	31	31	31	31
Number of Countries	16	16	16	16	16	16
R-Squared	0.383	0.169	0.155	0.532	0.474	0.570

Table A8: Bias in Economic Sentiment and Coverage, Beginnings of Articles. Newspaper fixed effects were included in the analysis, but omitted from the table. Standard errors clustered by newspaper are in parentheses. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interaction with the economy are jointly zero. $^{+}p < .10,^{*}p < .05,^{**}p < .01,^{***}p < .001$.

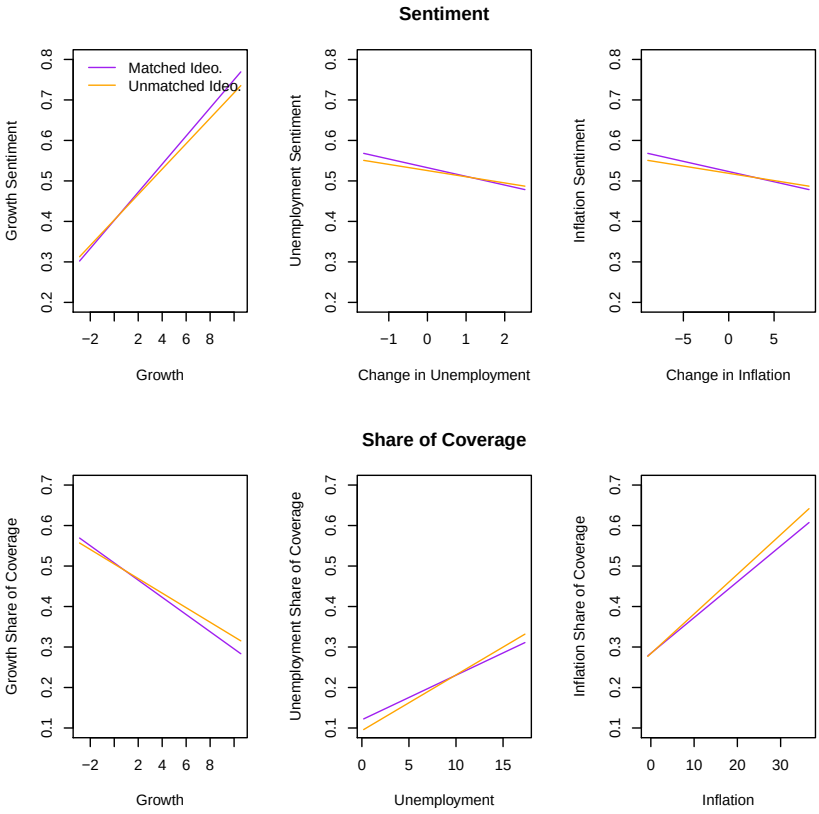

Figure A4: Sentiment and Coverage vs. the Economy for Ideologically Matched and Unmatched Newspapers, Beginnings of Articles. Results are calculated based on Table A7

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)
p-Value from Wald Test:						
Interactions between Government and Newspaper Ideology = 0	0.826	0.766	0.309	0.911	0.361	0.983

Table A9: Bias in Economic Sentiment and Coverage, Continuous Measures of Ideology. Newspaper fixed effects were included in the analysis. Standard errors were clustered by newspaper. In all cases, the null hypothesis in the Wald test is that the coefficient on Newspaper Ideology * Government Ideology and it's interaction with the economy are jointly zero. $^{+}p < .10,^{*}\, p < .05,^{**}\, p < .01,^{***}\, p < .001$.

Country	Newspaper	Relative Ideological Orientation
Australia	*The Age*	Center-Left (-1)
Australia	*Herald Sun*	Center-Right (1)
Austria	*Der Standard*	Left (-2)
Austria	*Die Presse*	Center-Right (1)
Canada	*Toronto Star*	Center-Left (-1)
Canada	*The Globe and Mail*	Center (0)
France	*Le Monde*	Center-Left (-1)
France	*Le Figaro*	Center-Right (1)
Germany	*Die Zeit*	Center-Left (-1)
Germany	*Frankfurter Allgemeine*	Center-Right (1)
Ireland	*Irish Times*	Left (-2)
Ireland	*Irish Independent*	Right (2)
Israel	*Globes*	Center (0)
Israel	*Jerusalem Post*	Right (2)
Italy	*La Stampa*	Center-Left (-1)
Italy	*Corriere della Sera*	Center-Right (1)
Japan	*Nikkei Weekly*	Center-Left (-1)
Japan	*Daily Yomiuri*	Right (2)
Luxembourg	*Le Quotidien*	Left (-2)
Luxembourg	*Le FAX d'Agefi*	Center (0)
New Zealand	*The Press*	Center-Left (-1)
New Zealand	*New Zealand Herald*	Right (2)
Portugal	*Correio da Manha*	Center (0)
Portugal	*Journal de Noticias*	Center-Right (1)
Spain	*El Pais*	Center-Left (-1)
Spain	*El Mundo*	Center-Right (1)
Switzerland	*Tages-Anzeiger*	Center-Left (-1)
Switzerland	*Neue Zurcher Zeitung*	Right (2)
United Kingdom	*The Guardian*	Left (-2)
United Kingdom	*London Times*	Right (2)
United States	*New York Times*	Center-Left (-1)
United States	*The Wall Street*	Center-Right (1)

Table A10: Five-point Ideological Orientation of Newspapers. Sources used include *eurotopics*, *nationsencyclopedia*, *propertyguides*, *voxeurop*, *wikipedia*, *world-newspapers*, and *worldpress*.

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)
p-Value from Wald Test:						
Ideology Terms = 0	0.977	0.443	0.579	0.232	0.175	0.296

Table A11: Bias in Economic Sentiment and Coverage, Left vs. Right-wing Papers. Newspaper fixed effects were included in the analysis. Standard errors were clustered by newspaper. In all cases, the null hypothesis in the Wald test is that the coefficient on Ideological Match and it's interactions are jointly zero. $^{+}p < .10,^{*}p < .05,^{**}p < .01,^{***}p < .001$.

Dependent Variable:	Economic Sentiment (articles)	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)
Independent Variables:				
Constant	0.463***	0.433***	0.513***	0.510***
	(0.008)	(0.012)	(0.005)	(0.004)
Growth (yearly)	0.014***	0.024***	0.011***	-0.003+
(SD = 3.053)	(0.002)	(0.004)	(0.002)	(0.001)
Change in Unem. (yearly)	-0.008	-0.022+	-0.014***	0.007*
(SD = 0.972)	(0.005)	(0.009)	(0.003)	(0.003)
Change in Inf. (yearly)	-0.003+	-0.004	0.002	-0.008***
(SD = 5.561)	(0.001)	(0.002)	(0.001)	(0.002)
Number of Months	6648	6647	6638	6581
Number of Newspapers	32	32	32	32
Number of Countries	16	16	16	16
R-Squared	0.468	0.482	0.300	0.249

Table A12: The Effect of the Economy on Newspaper Sentiment (bootstrapped p-values). Newspaper fixed effects were included in each regression, but omitted from the table. Standard errors clustered by newspaper in parentheses. P-values were computed using the wild bootstrap. $^{+}p < .10,^{*}p < .05,^{**}p < .01,^{***}p < .001$.

Dependent Variable:	Economic Coverage (articles)	Growth Share of Coverage (articles)	Unemployment Share of Coverage (articles)	Inflation Share of Coverage (articles)
Independent Variables:				
Growth (yearly)	-0.022**	-0.016***		
	(0.006)	(0.002)		
Unemployment (yearly)	0.002		0.012***	
	(0.005)		(0.002)	
Inflation (yearly)	-0.008			0.008+
	(0.007)			(0.003)
Number of Months	6136	6648	6656	6660
Number of Newspapers	32	32	32	32
Number of Countries	16	16	16	16
R-Squared	0.775	0.640	0.630	0.698

Table A13: The Effect of the Economy on Newspaper Coverage (bootstrapped p-values). Newspaper fixed effects were included in each equation, but omitted from the table. Standard errors clustered by newspaper in parentheses. P-values were computed using the wild bootstrap. $^{+}p < .10$,* $p < .05$,** $p < .01$,*** $p < .001$.

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Sentiment (headlines)	Unemployment Sentiment (headlines)	Inflation Sentiment (headlines)
Independent Variables:						
Growth (yearly)	0.027***	0.010***	-0.003+	0.033***	0.017*	-0.007*
	(0.005)	(0.002)	(0.001)	(0.006)	(0.006)	(0.002)
Change in Unemployment (yearly)		-0.013*			-0.011	
		(0.005)			(0.009)	
Change in Inflation (yearly)			-0.010**			-0.013**
			(0.002)			(0.004)
Ideological Match	-0.001	-0.002	0.003	0.016	0.007	-0.005
	(0.019)	(0.009)	(0.007)	(0.018)	(0.021)	(0.014)
Growth * Ideo. Match	0.003	0.002	-0.003	0.002	0.002	0.004
	(0.006)	(0.003)	(0.002)	(0.007)	(0.007)	(0.003)
Change in Unem. * Ideo. Match		-0.004			0.004	
		(0.006)			(0.011)	
Change in Inf. * Ideo. Match			0.003			0.003
			(0.003)			(0.004)
p-Value from Wald Test:						
Ideology Terms = 0	0.872	0.899	0.353	0.127	0.830	0.705
Number of Months	6512	6504	6448	5959	5243	5564
Number of Newspapers	32	32	32	31	31	31
Number of Countries	16	16	16	16	16	16
R-Squared	0.462	0.302	0.244	0.171	0.063	0.055

Table A14: Media Bias in Sentiment (bootstrapped p-values). Newspaper fixed effects were included in the analysis, but omitted from the table. Standard errors clustered by newspaper in parentheses. P-values were computed using the wild bootstrap. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interactions with the economy are jointly zero. $^{+}p < .10,^{*}\, p < .05,^{**}\, p < .01,^{***}\, p < .001$.

Dependent Variable:	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)	Growth Coverage (headlines)	Unemployment Coverage (headlines)	Inflation Coverage (headlines)
Independent Variables:						
Growth (yearly)	-0.015***			-0.015***		
	(0.002)			(0.003)		
Unemployment (yearly)		0.013***			0.015***	
		(0.002)			(0.002)	
Inflation (yearly)			0.010***			0.009***
			(0.002)			(0.003)
Ideological Match	0.003	0.023	0.006	0.002	-0.004	0.008
	(0.010)	(0.017)	(0.017)	(0.012)	(0.026)	(0.022)
Growth * Ideo. Match	-0.003			-0.005		
	(0.003)			(0.004)		
Unem. * Ideo. Match		-0.003+			0.000	
		(0.002)			(0.003)	
Inf. * Ideo. Match			-0.002			0.002
			(0.004)			(0.004)
p-Value from Wald Test:						
Ideology Terms = 0	0.523	0.184	0.937	0.544	0.881	0.869
Number of Months	6513	6521	6525	6269	6274	6276
Number of Newspapers	32	32	32	31	31	31
Number of Countries	16	16	16	16	16	16
R-Squared	0.641	0.626	0.697	0.305	0.226	0.308

Table A15: Media Bias in Coverage (bootstrapped p-values). Newspaper fixed effects were included in the analysis, but omitted from the table. Standard errors clustered by newspaper in parentheses. P-values were computed using the wild bootstrap. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interaction with the economy are jointly zero. $^{+}p < .10,^{*}\, p < .05,^{**}\, p < .01,^{***}\, p < .001$.

Dependent Variable:	Growth Sentiment (articles)	Unemployment Sentiment (articles)	Inflation Sentiment (articles)	Growth Coverage (articles)	Unemployment Coverage (articles)	Inflation Coverage (articles)
Independent Variables:						
Growth (yearly)	0.118**	0.100	0.201	-0.061**		
	(0.044)	(0.123)	(0.202)	(0.022)		
Unemployment (yearly)					0.132***	
					(0.035)	
Change in Unemployment (yearly)		-0.826**				
		(0.273)				
Inflation (yearly)						0.185***
						(0.040)
Change in Inflation (yearly)			-0.192			
			(0.181)			
Ideological Match	-0.044	0.197	1.198	-0.052	0.390	0.423
	(0.284)	(0.481)	(1.961)	(0.183)	(0.278)	(0.362)
Growth * Ideo. Match	0.059	-0.041	-0.198	-0.004		
	(0.085)	(0.174)	(0.571)	(0.055)		
Unem. * Ideo. Match					-0.032	
					(0.034)	
Change in Unem. * Ideo. Match		0.714				
		(0.435)				
Inf. * Ideo. Match						-0.242***
						(0.060)
Change in Inf. * Ideo. Match			0.155			
			(0.252)			
p-Value from Wald Test:						
Ideology Terms = 0	0.688	0.254	0.700	0.882	0.310	0.000***
Number of Months	883	319	137	2093	2003	1987
Number of Newspapers	25	23	20	26	25	25
Number of Countries	13	13	11	13	13	13

Table A16: Bias in Human-coded Sentiment and Coverage. Newspaper fixed effects were included in the analysis, but omitted from the table. Standard errors clustered by newspaper are in parentheses. In all cases, the null hypothesis in the Wald test is that Ideological Match and it's interaction with the economy are jointly zero. $^{+}p < .10$,$^{*}p < .05$,$^{**}p < .01$,$^{***}p < .001$.

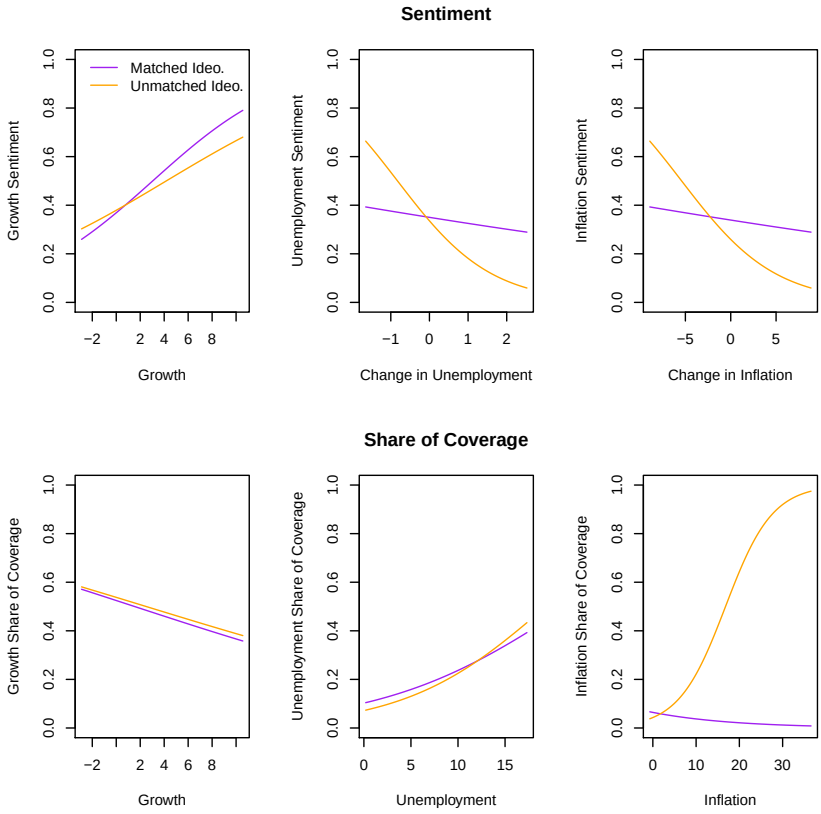

Figure A5: Human-coded Sentiment and Share of Coverage vs. the Economy for Ideologically Matched and Unmatched Newspapers. Results are calculated based on Table A16

References

Cohen, Jacob. 1960. "A Coefficient of Agreement for Nominal Scales." *Educational and Psychological Measurement* 20:37–46.