Appendix – How Newspapers Reveal Political Power

Pamela Ban*
Harvard University

Alexander Fouirnaies†
Harris School, University of Chicago

Andrew B. Hall‡
Stanford University

James M. Snyder, Jr.§
Harvard University and NBER

December 6, 2017

*Corresponding Author. Ph.D. Candidate. pban@fas.harvard.edu.
†Assistant Professor. fouirnaies@uchicago.edu.
‡Assistant Professor. andrewbhall@stanford.edu.
§Professor. jsnyder@gov.harvard.edu.

A. 1 Information on Dataset

In this section, we describe in detail the dataset on newspapers that we have compiled.

A. 1.1 Summary of Data

The dataset consists of a stratified sample of pages printed in U.S. local newspapers initially published during the period 1877-1977 and later reproduced by Newspapers.com. The stratification works as follows. As new pages are added to the dataset, they are assigned a number based on the newspaper issue in which they belong. We sample all pages that end with the integers 1, 2, ..., 7. At any given moment in time, we thus sample 70% of the existing newspaper data, but because new data is added all the time, we cannot offer a precise percentage for future dates.

Each page in our dataset is a string of characters and spaces extracted by Newspapers.com from scanned copies of the original newspaper pages using OCR techniques, and each page is connected to the following meta data: name of newspaper, publication date, page number, state, county and city of publication. In total, the dataset contains approximately 50 million unique pages from 2700 newspapers distributed across approximately a thousand counties in the US.

A. 1.2 Geographical Coverage

Using this metadata we count the number of pages and newspapers in each state and report this in the map in Figure A.1. The dataset geographically covers all states, and approximately a third of all counties appear in the dataset at some point during the studied period. The dataset roughly reflects the population density over the studied period. The states that most frequently appear in the dataset are PA, TX, CA, IL, OH (ranging from 3-6 million pages), whereas less populated states such as WY, ME, RI, VT each contribute with approximately 100,000 pages. Table A.1 reports the exact number of pages for each state.

A. 1.3 Temporal Coverage

Using the publication dates obtained from the meta data, we count the total number of pages published each year in the four Census regions. These numbers are reported in Figure A.2. The graph illustrates that the number of newspaper pages increase over the first 30-40 years, then stagnates until the late 1940s and then rapidly grow over the rest of the studied period. The temporal patterns are fairly consistent across regions.

Figure A.1 – Geographical Distribution of Pages and Newspapers in Dataset. Darker shaded areas reflect more pages. The digits on the map report the total number of unique newspapers in the state that appear in the sample.

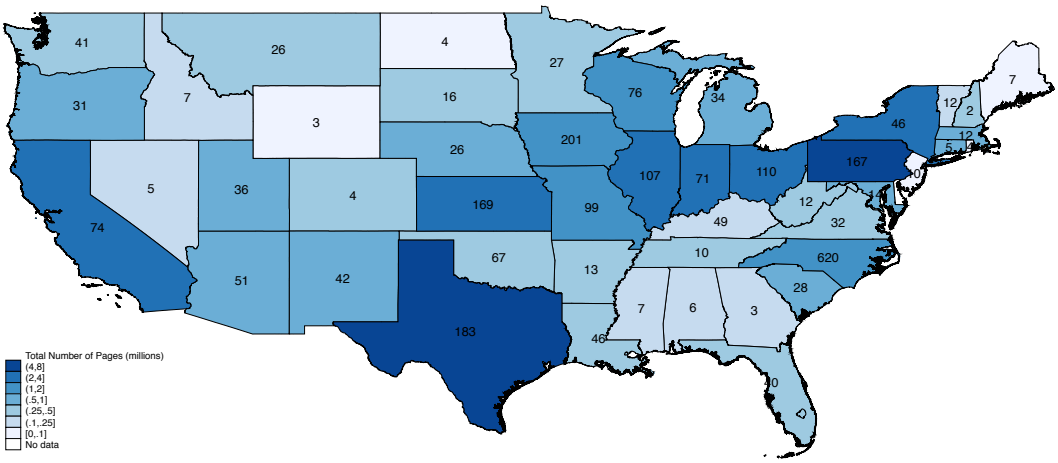


Figure A.2 – Yearly Number of Pages in Sample by Region.

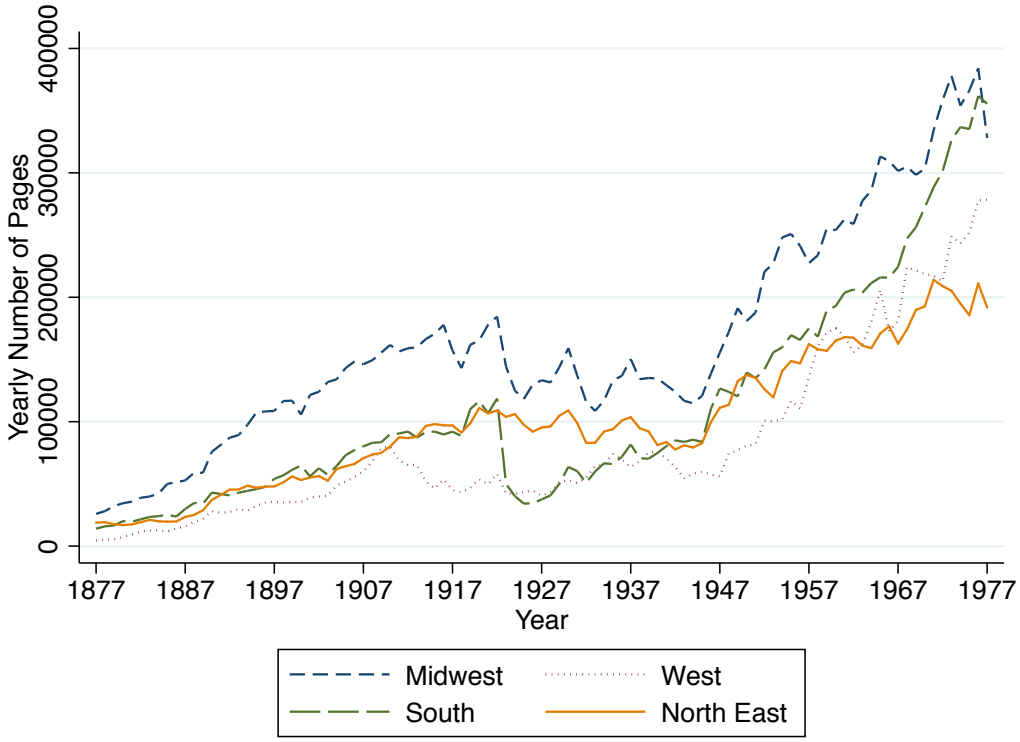


Table A.1 – Number of Pages, Newspapers and Counties in Dataset by State

	1877-1902			1903-1927			1928-1952			1953-1977			Total		
	Pages (1000s)	Papers	Counties	Pages (1000s)	Papers	Counties	Pages (1000s)	Papers	Counties	Pages (1000s)	Papers	Counties	Pages (1M)	Papers	Counties
AK	0.0	1	1	4.0	5	2	17.4	3	3	80.9	2	2	0.1	8	4
AL	0.0	1	1	4.4	1	1	66.3	4	3	127.5	2	2	0.2	6	4
AR	0.4	2	1	4.5	2	1	83.9	8	5	208.2	9	5	0.3	13	6
AZ	40.8	29	11	92.6	23	11	96.1	9	7	443.0	11	7	0.7	51	12
CA	244.0	25	12	607.5	34	15	653.9	28	16	2136.7	43	18	3.6	74	24
CO	4.9	2	2	2.9	2	2	40.1	1	1	251.1	3	2	0.3	4	3
CT	3.1	1	1	33.5	1	1	25.2	3	2	480.3	5	3	0.5	5	3
DC	180.1	17	1	449.1	10	1	0.0	0	0	0.0	0	0	0.6	21	1
FL	5.2	12	11	33.4	34	19	16.4	3	2	216.1	5	5	0.3	40	21
GA	56.8	2	2	90.8	2	1	0.0	0	0	0.0	0	0	0.1	3	2
HI	49.0	13	3	68.8	10	4	0.0	0	0	0.0	0	0	0.1	16	4
IA	100.9	79	29	135.3	40	18	379.1	104	61	628.7	114	65	1.2	201	78
ID	12.9	1	1	0.4	1	1	48.1	5	3	131.0	3	3	0.2	7	5
IL	323.3	47	19	579.0	48	15	599.0	45	18	1338.3	56	19	2.8	107	30
IN	249.6	37	15	657.3	37	16	480.0	29	20	981.2	31	17	2.4	71	23
KS	576.8	133	50	944.4	100	48	122.3	13	11	446.9	13	12	2.1	169	58
KY	109.4	35	25	101.3	39	28	16.0	1	1	17.8	3	2	0.2	49	31
LA	80.7	30	21	42.0	28	21	60.4	4	2	167.3	4	3	0.4	46	27
MA	55.5	9	4	82.5	7	3	130.7	5	3	248.6	6	3	0.5	12	4
MD	23.0	8	5	43.7	8	6	238.8	9	6	627.6	11	7	0.9	14	7
ME	15.4	2	1	0.0	0	0	10.0	2	1	28.4	3	2	0.1	7	4
MI	20.5	21	5	54.3	9	4	233.1	15	10	511.7	14	9	0.8	34	11
MN	79.9	19	14	93.6	17	13	24.2	4	4	148.0	8	5	0.3	27	17
MO	136.6	49	26	345.1	67	27	448.0	39	17	865.2	27	14	1.8	99	33
MS	0.0	0	0	5.3	4	2	34.1	4	2	83.7	1	1	0.1	7	2
MT	21.2	11	7	70.5	13	12	151.6	9	6	255.2	8	6	0.5	26	16
NC	388.4	490	80	845.8	234	75	149.4	36	24	415.8	12	10	1.8	620	84
ND	18.6	4	1	41.6	2	1	28.7	1	1	8.9	1	1	0.1	4	1
NE	52.5	19	13	137.5	19	14	214.1	6	3	305.1	5	3	0.7	26	16
NH	5.2	1	1	28.4	1	1	48.5	2	2	177.7	2	2	0.3	2	2
NJ	16.7	8	4	47.1	4	3	7.8	2	2	16.9	3	3	0.1	10	5
NM	28.6	19	12	38.5	23	15	135.8	9	9	556.6	16	13	0.8	42	23
NV	18.7	3	1	30.5	2	1	51.8	2	1	54.4	3	1	0.2	5	1
NY	291.4	22	10	660.6	24	11	549.2	22	15	701.6	18	13	2.2	46	17
OH	230.9	63	32	524.6	58	26	762.4	37	23	1262.5	47	28	2.8	110	42
OK	29.3	25	12	169.6	35	15	72.9	22	9	163.2	13	7	0.4	67	21
OR	49.4	16	11	205.2	26	15	199.8	8	5	105.2	6	5	0.6	31	16
PA	481.2	101	50	1332.0	89	42	1705.2	73	36	2547.3	69	33	6.1	167	55
RI	8.4	2	1	9.1	2	1	16.8	3	1	64.8	3	1	0.1	4	1
SC	25.3	22	15	88.5	19	13	127.0	5	4	268.1	4	4	0.5	28	16
SD	28.6	10	4	66.7	8	3	71.7	8	5	152.1	6	4	0.3	16	5
TN	3.5	6	5	6.9	1	1	69.9	4	1	198.2	3	1	0.3	10	5
TX	143.8	44	30	472.5	82	45	1100.2	85	53	2917.8	106	61	4.6	183	72
UT	94.8	18	7	169.1	28	10	231.3	4	3	460.1	5	4	1.0	36	11
VA	97.3	25	18	51.2	21	17	49.4	2	1	217.3	4	3	0.4	32	22
VT	22.0	12	9	12.1	6	6	0.0	0	0	67.0	1	1	0.1	12	9
WA	15.2	16	10	53.3	29	16	16.8	5	4	217.0	5	3	0.3	41	21
WI	70.9	38	20	221.2	27	20	226.6	29	18	708.3	21	15	1.2	76	32
WV	0.9	2	1	20.4	6	4	58.2	4	3	310.3	8	5	0.4	12	6
WY	0.0	0	0	0.8	3	3	0.9	1	1	1.5	1	1	0.0	3	3
All States	4511.9	1552	615	9779.6	1291	630	9869.1	717	428	22320.9	744	434	46.5	2700	916

Table A.2 – The 50 Most Common Newspapers in Dataset.

Newspaper	Pages	First Year	Last Year	State
Abilene Reporter-News	452,252	1926	1977	TX
Albuquerque Journal	312,826	1882	1977	NM
Alton Evening Telegraph	235,141	1853	1972	IL
The Bridgeport Post	273,910	1947	1977	CT
The Bridgeport Telegram	227,785	1918	1977	CT
The Brooklyn Daily Eagle	457,294	1841	1955	NY
Chicago Daily Tribune	257,688	1849	1922	IL
The Chillicothe Constitution-Tribune	224,239	1890	1988	MO
The Cincinnati Enquirer	195,487	1841	1923	OH
The Corpus Christi Caller-Times	241,515	1912	1977	TX
The Daily Herald	429,998	1886	2006	UT
The Daily Times	205,312	1865	1977	NJ
Delaware County Daily Times	286,222	1876	1977	IN
El Paso Herald-Post	193,431	1931	1977	TX
The Evening News	194,214	1899	1974	MI
The Evening Review	231,344	1885	1977	OH
The Galveston Daily News	319,238	1865	1999	TX
The Gettysburg Times	213,953	1909	2009	PA
The Index-Journal	396,147	1919	2010	SC
The Indiana Gazette	323,554	1868	1981	PA
Indiana Gazette	201,415	1890	2008	PA
The Indianapolis News	193,653	1869	1932	IN
The Kansas City Star	340,728	1881	1976	MO
The Kokomo Tribune	347,354	1868	1999	IN
Lebanon Daily News	247,459	1872	1977	PA
Lincoln Evening Journal	230,925	1912	1976	NE
The Lincoln Star	300,099	1913	1977	NE
Logansport Pharos-Tribune	205,433	1890	2006	IN
Lubbock Avalanche-Journal	316,812	1927	1977	TX
The Morning Herald	427,066	1907	1977	MD
New Castle News	363,846	1891	1978	PA
The New York Times	259,388	1851	1922	NY
News-Journal	198,110	1891	1977	OH
The News-Palladium	229,649	1896	1978	MI
The Ogden Standard-Examiner	309,659	1888	1977	UT
The Oil City Derrick	201,981	1885	1977	PA
Oshkosh Daily Northwestern	219,797	1872	1975	WI
The Ottawa Journal	510,633	1885	1980	PA
The Pantagraph	250,388	1954	2013	IL
The Paris News	237,867	1933	1999	TX
The Post-Crescent	195,471	1861	1976	WI
The Salina Journal	287,177	1951	2009	KS
The Salt Lake Tribune	334,311	1890	1977	UT
The San Bernardino County Sun	698,155	1894	1998	CA
Santa Ana Register	214,518	1906	1977	CA
Santa Cruz Sentinel	482,474	1884	2005	CA
The Sedalia Democrat	219,671	1891	1987	MO
Standard-Speaker	232,882	1961	2000	PA
The Times	742,550	1785	1998	NY
Tucson Daily Citizen	234,102	1941	1977	AZ

A. 1.4 Commands Used to Process Text

It is impossible to extract large amounts of text from old newspapers without any errors. Smeared ink, pictures, poor paper quality, variation in font types, dirty scanners as well as typos in the original articles are among the sources of errors. Common errors are: the letter "c" is read but the actual letter is "e" and vice-versa; "a" vs. "u"; "t" vs. "l" vs. "i"; "g" vs. "q" vs. "y"; and "m" vs. "rn." Hyphenation is also a serious issue—since newspaper columns are narrow many words must be split and hyphenated. Extra spaces and stray marks are also common.

Most of these errors will be random and add noise to the word counts. To reduce this noise, we follow the common approach of using regular expressions when we search for words. We carefully read through a large number of newspaper pages and compared the OCR text with the original newspaper page. Based on this material, we identified a number of common errors and use the regular expressions outlined in Table A.3 to catch these errors. Before searching in the string, we substitute all upper case characters to lower case.

Table A.3 – Regular Expressions.

Error Type	Correct Character	OCR	Regex	Example
1:1 Substitution	e	c	[ec]	s[ec]nate
	v	y	[vy]	executi[vy]e
	o	c	[oc]	c[oc]mmittee
	i	l	[il]	comm[il]ttee
	t	l	[tl]	commi[tl][tl]ee
	b	h	[bh]	[bh]udget
	g	y,j,q	[gyjq]	bud[gyjq]et
	f	t	[ft]	o[ft][ft]ice
	a	u,o	[auo]	sen[auo]te
1:2 Substitution	m	rm	[m(rn)]	co[m(rn)][m(rn)]ittee

A. 2 Additional Analyses

In this section we offer follow-up analyses and robustness checks to extend the estimates presented in the paper.

A. 2.1 The Reciprocal Trade Agreement Act

In 1934 Congress passed and President Roosevelt signed the Reciprocal Trade Agreement Act (RTAA). This law gave the President the authority to negotiate reciprocal tariff agreements with other nations. These agreements could increase or a decrease import duties by up to 50 percent, and did not require congressional approval.

There is widespread agreement that this act represented a substantial transfer of power over tariff policy, from Congress to the President. For example, Haggard (1988: 112) writes that in passing the RTAA "the most important issues at stake in 1934 were institutional, centering on the transfer of authority from Congress to the executive." Irwin (1998: 325) writes: "From the Civil War up to the Smoot-Hawley tariff of 1930, Congress retained exclusive authority over U.S. tariffs, which for the most part consisted of a single-column schedule of nonnegotiable, nondiscriminatory import duties... [With the RTAA], Congress granted the president the authority to reach tariff reduction agreements—agreements that did not require congressional approval—with foreign countries." Kaplan (1996: 45) writes: "the RTAA Act would significantly reduce the power of Congress in the tariff-making process."[1,2]

As another check on the idea that media coverage can be used to measure power, we examine whether coverage of tariff policymaking shifted away from Congress and toward the President after the passage of the RTAA. More specifically, to measure the coverage of Congress in tariff policymaking we include all cases where "congress" or "house" or "senate" appeared within five words of "tariff"—call this *Congress*. Similarly, to measure the coverage of the President in tariff policymaking we include all cases where "president" or "administration" appeared within five words of "tariff"—call this *President*. We then make the share of coverage devoted to Congress in each time period t:

$$\textit{Relative Coverage of Congress}_t = \frac{\textit{Congress}_t}{\textit{Congress}_t + \textit{President}_t}.$$

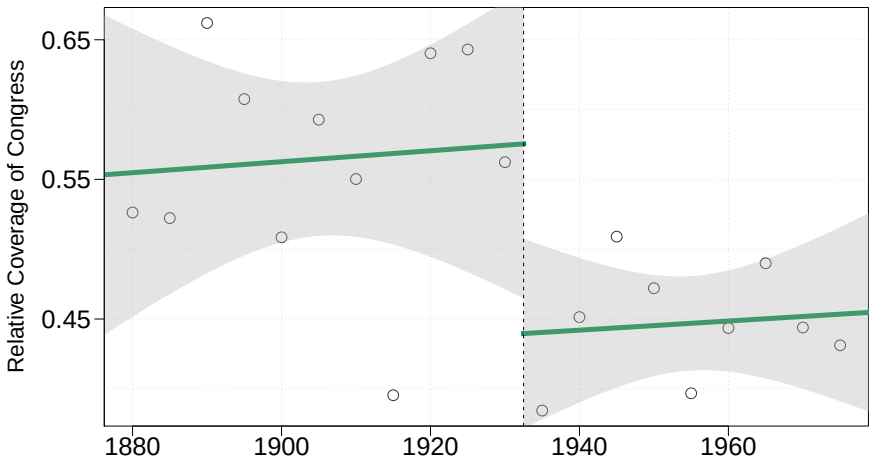
Figure A.3 shows a graph of *Relative Coverage of Congress* over time. We average over 5-year periods, so the point labeled 1930 covers the years 1930-1934, the point labeled 1935 covers 1935-1939, etc. The figure shows clearly that newspaper coverage of Congress relative to the President fell sharply after 1934. Before the RTAA Congress had about about 55% of the mentions, while after the RTAA this fell to only about 40% of the mentions. This is what we expect if relative newspaper coverage is a reasonable proxy for the relative power of the two branches over tariff policy.

The outlier in the pre-1935 period, covering the years 1915-1919, covers the years in which the U.S. was directly involved in WWI and during which the Wilson administration fought for the

[1]For more such quotes, see: Shoch (2001: 56); Schnietz (2000: 417); Bordreaux (2008: 121); and Irwin (2009: 221).

[2]Congress did not cede permanent authority to negotiate tariffs to the President, but set the RTAA to expire every three years or less. However, as many scholars point out, extending the RTAA was quite different than passing bills containing the entire schedule of tariffs for all imported goods across the entire country. The RTAA was renewed in 1937, 1940, 1943, 1945, 1948, 1949, 1951, 1953, 1954, 1955, 1958. In 1962 Congress passed the Trade Expansion Act of 1962, granting the President authority for five years to enter into agreements that negotiated the reduction or elimination of tariffs. That act also expanded Congress's role in the negotiating process, by requiring the President to submit for congressional review a copy of each concluded agreement and a presidential statement explaining why the agreement was necessary." See, e.g., Fergusson (2015) and Bailey, Goldstein, and Weingast (1997).

Figure A.3 – Relative Coverage of Congress in Tariff Policymaking. The measured power of Congress in the realm of tariff policy decreased abruptly after the passage of the RTAA.



League of Nations. It is possible that these events contributed to the exceptionally high relative coverage of the president during this period. Finally, we should note that an OLS regression shows that the change is highly significant statistically as well as substantively.

A. 2.2 Additional Details on Congressional Party Leaders

In Table A.4, we report the average yearly number of hits five years before, during and five years after the leadership period.[3] Two things are worth noting. First, similar to the results presented in the main text, Panel A shows that on average the news coverage of members of Congress increases by an order of magnitude when they serve as Speakers. Second, we see a similar pattern for minority-party leaders.[4] When a member of Congress is appointed to leader of the minority party, the member receives more coverage in the newspapers. However, the media boost for minority-party leaders is not quite as big as the boost enjoyed by Speakers. This difference probably reflects that Speakers are more powerful than minority-party leaders. Overall, the results presented in Table A.4 further supports the idea that power is reflected in the newspaper coverage.

A. 2.3 Additional Details on Congressional Committees

As discussed in the main text, our coverage-based power measure, when applied to Congressional committees, is highly correlated with the Groseclose-Stewart ranking. There are a few outliers, which we believe go in our favor. For example, the Committee on House Administration is ranked higher in the Groseclose–Stewart ranking than in our coverage-based ranking. House Administration is probably quite weak rather than powerful, in the sense that its jurisdiction, revenue-raising ability, and influence over policy outcome is limited, though it may be more "desirable" to members of the House since, after all, it deals with House matters (and people care about themselves).

[3]For the party leaders who served in several non-consecutive periods, we classify the hits from the "middle" period (when they were not in power) as belonging to the post-leadership period. None of the results are sensitive to this classification.

[4]We only include minority-party leaders who did not serve as Speaker five years before and after he served as minority-party leader.

Table A.4 – News Coverage Before, During and After Leadership Term. Serving as party leader substantially increases the news coverage of members of Congress.

	Panel A: Speakers		
	Before	During	After
Hits	42.94	315.31	42.92
	(61.00)	(502.41)	(88.32)
Difference	-272.37		-272.39
P-value	0.00		0.00
N	86	113	98
	Panel B: Minority Leaders		
Hits	23.75	139.67	47.74
	(31.96)	(106.14)	(75.72)
Difference	-115.92		-91.93
P-value	0.00		0.00
N	20	30	23

Standard deviations are reported in parentheses. The pre and post-Speaker periods are based on 5 years before and after the Speaker term.

Another outlier worth mentioning is the Judiciary Committee. Judiciary ranks high based on our coverage-based measure, but ranks near the middle in the Groseclose-Stewart ranking. A possible contributor to this divergence is Watergate, which was highly covered in the press. This represents one of the limitations of our measure – since our measure is based on relative press coverage, any "sensational" event that temporarily increases press coverage of a political actor or group even though the underlying power of that actor or group remains the same would result in measurement error. In the case of Judiciary, was the increase in coverage exclusively due to the sensational nature of Watergate, or did the Judiciary Committee at that time truly hold a significant increase in the amount of power, since they were presented with a rare instance in which they could use their power over the impeachment of a president? This example reflects the need to carefully apply our measure and consider possible explanations for sharp fluctuations.

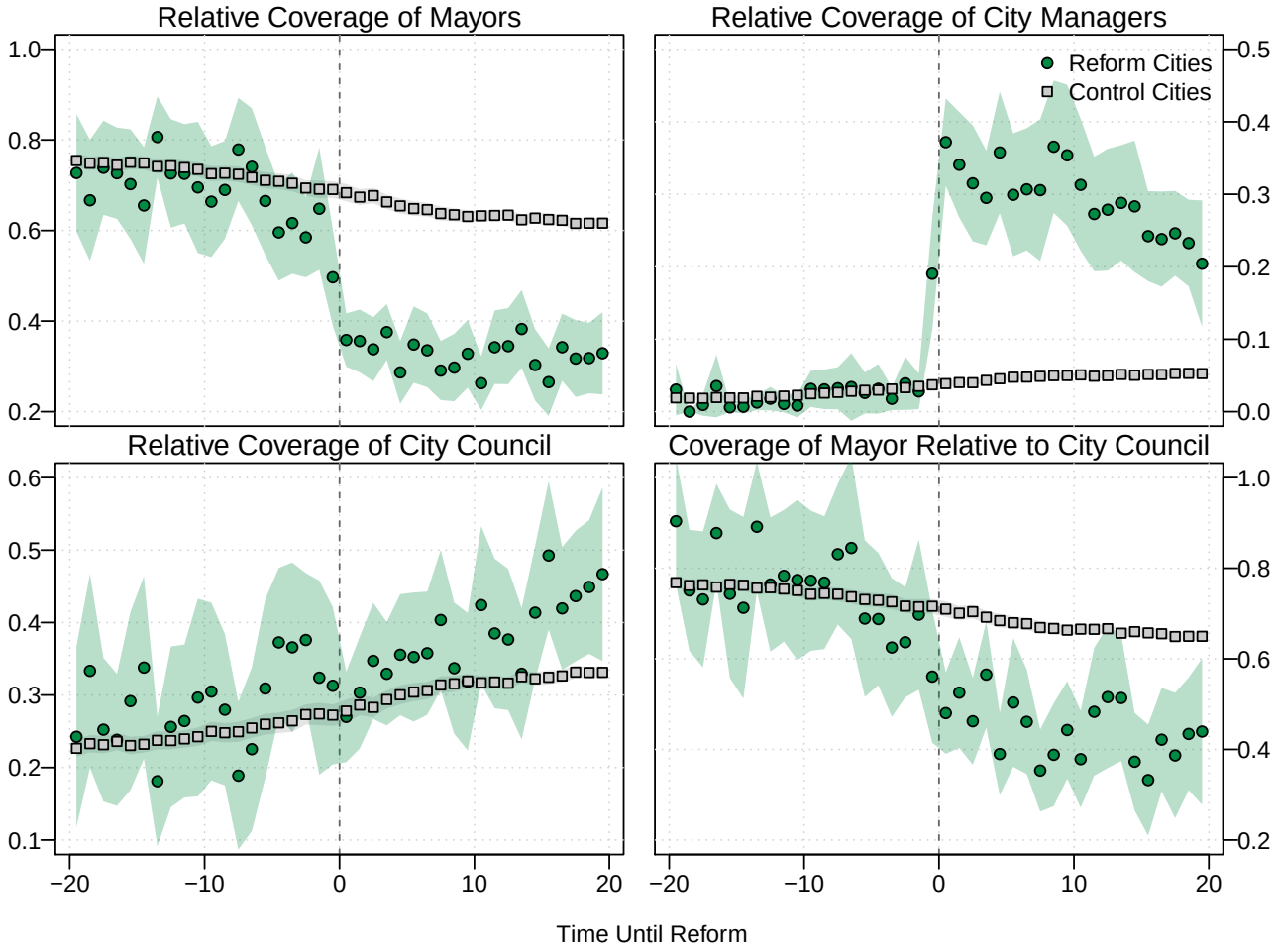
A. 2.4 Additional Results on Mayoral Reforms

In this subsection, we also perform two additional analyses on the effects of city reforms that stripped the mayor of powers and reallocated them to the city manager. First, in Figure A.4, we replicate the figure from the body of the paper but employing city name filtering. Specifically, we limit the mentions of the word "mayor" to only those that appear near the mention of the mayor's home city. This removes false positives that occur when newspapers discuss *other* cities' mayors.[5]

[5]Note that this misses a large number of "correct" mentions. For example, newspapers often give the name of the mayor or city manager near the relevant search string, rather than the name of the city. A better idea is to limit attention to mentions in which the name of the newspaper's home city or the name of the mayor (or city manager)

As the plot shows, we continue to find the same pattern of results; in fact, if anything, the decrease in the coverage of mayors and the increase in the coverage of city managers is even more pronounced than before.

Figure A.4 – Relative Coverage of City Offices Over Time: Filtering Results by City Name. Here we replicate the analysis from the main text, but we filter mentions of mayors to only include those where the name of the mayor's city is mentioned nearby in the text. Again, city government reforms are seen to reduce the measured power of mayors and increase that of city managers and city council members.



Second, we also re-run the formal difference-in-differences estimation from the main text. The difference-in-differences relies on the so-called "parallel trends" assumption. Here, we assess the robustness of our results by relaxing this assumption. Specifically, we include linear, city-specific time trends. Table A.5 displays the results. As it shows, the results are nearly identical to those in the main text.

appears near the relevant search string. This, however, requires lists of all of the mayors serving during the relevant time periods for all cities in our sample. We are currently compiling these lists, but do not have them yet.

Table A.5 – Impact of Switch from Mayor-Council to Council-Manager City Government. Results from a difference-in-differences design suggest that the reform causes a large decrease in the relative coverage of mayors.

	All Mentions		Using City Name Filter	
	Relative Coverage of Mayor	Relative Coverage of City Manager	Relative Coverage of Mayor	Relative Coverage of City Manager
Council-Manager Govt Form	-0.18	0.20	-0.27	0.31
	(0.03)	(0.03)	(0.04)	(0.04)
N	3540	3540	2376	2376
City Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
City-Specific Time Trends	Yes	Yes	Yes	Yes

Standard errors, clustered by city, are in parentheses.

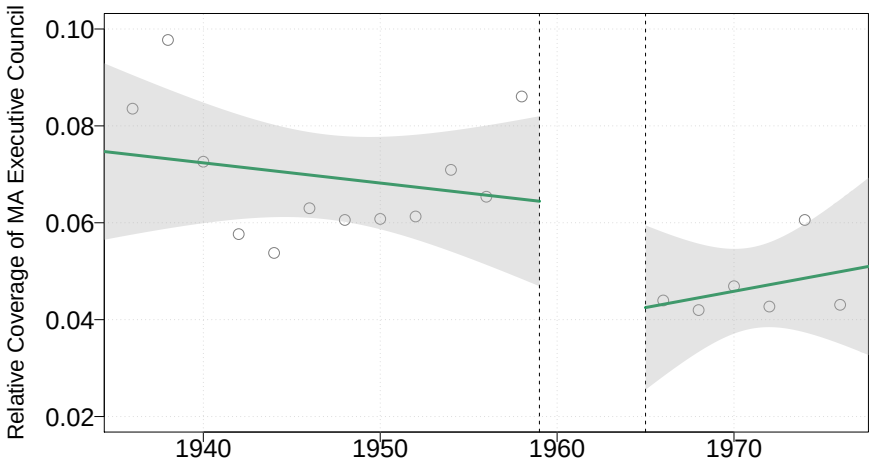
A. 2.5 Additional Results on MA Council Reform

In the main text, we showed how the coverage of the MA executive council changed after a reform stripping it of many of its powers. In that figure, we used all available newspaper data. Now, we replicate the analysis but only using the Boston Globe, to make sure the results are not driven by our dataset. We thus re-calculate our relative coverage measure using only mentions in the Boston Globe. Figure A.5 presents the results. We continue to see a sharp drop after the reform.

A. 2.6 Correlation with Mayhew TPO Scores

On the basis of an exhaustive reading of secondary sources, Mayhew (1986) assigns "traditional party organization" (TPO) scores for each state on a scale from 1 (weak) to 5 (strong). As he notes, these scores are meant to capture the organizational strength "in the late 1960s" (Mayhew 1986: 6). If we consider the period 1966-1970, the correlation between *Party Mentions* and TPO is 0.56. If we focus just on the years 1968-1970 the correlation is even better, 0.63. This gives us some initial confidence in applying our measure to state party organizations.

Figure A.5 – Relative Coverage of the Massachusetts Executive Council Over Time: Boston Globe Coverage. The reform that stripped the Massachusetts Executive Council of its powers appears to decrease the coverage of the Executive Council relative to that of the Governor, who absorbed the power previously held by the council.



Note: The plot omits the years 1959–1965, during which discussion of the council spiked because of the scandal.