Editorial Influence on Newspaper Coverage:

American Newspapers 1869-1890

Work in Progress: Extended Essay

Winter Term 2025

EC 465

Economic Growth, Development, and Capitalism in Historical

Perspective

1 Introduction

I examine American newspapers in 1869-1890 and consider whether news coverage is meaningfully affected by changes in the ownership or editorship of the newspaper. To do this, I use editions of *Geo. P. Rowell and Co's American Newspaper Directory* to assemble a dataset of American newspapers with each paper's respective owners. I then observe news coverage using articles digitized in *American Stories*, and determine whether news coverage significantly changes in the years immediately following a change in the paper's editor or owner.

It might seem self-evident that the owner, and, in particular, the editor of a newspaper, have an affect on what topics the paper covers. But in the late nineteenth century, the entire business model of U.S. newspapers was upended as advertising revenue became the primary source of income, which drove dramatic changes in coverage as newspapers concerned themselves first and foremost with satiating consumer demand.[1] Furthermore, there is evidence that ideological slant in modern newspapers is driven far more by consumer demand than by the identity of the owners (Gentzkow and Shapiro, 2010), although more recent work on digital media adds nuance (Leung & Strumpf 2024). It is then interesting to consider the effects of owners and editors not on ideological slant but on what kind of news is covered (and what isn't).

To measure what news a newspaper covers, I construct counts of keywords in headlines gathered from the *American Stories* dataset, and associate each keyword with one of 10 different topics. Yearly counts per topic are made for every newspaper in the dataset with enough data to support measurement (roughly 600 different newspapers). Observations of changes in editor or publisher are then made for each of these newspapers by cross referencing with a dataset extracted from Rowell's Newspaper directories. The challenge then is to determine the degree to which these changes influence what news a paper decides to cover.

[1]See Baldasty 1992 or Petrova 2011

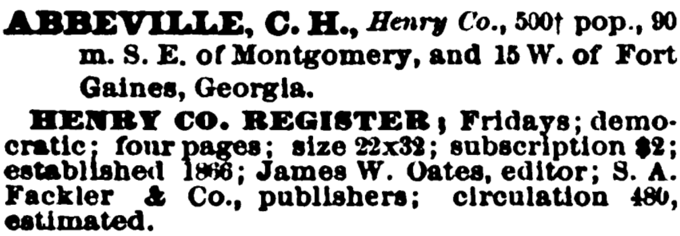
I employ a Difference-in-Differences design using a multidimensional distance metric as the dependent variable. Specifically, I calculate the Euclidean distance between a newspaper's post-treatment topic distribution and its pre-treatment baseline vector. This approach allows me to capture "structural drift" in coverage composition that is orthogonal to specific directional shifts (e.g., a paper could drift by becoming either more political or more commercial).

This essay proceeds as follows. In section 2, I discuss in more detail the data used to conduct this study. Section 3 briefly establishes the methodology behind the model used to estimate the effects of publishers and editors. Section 4 reports the results of my findings, and section 5 concludes the essay with some discussion of the limitations of these results and avenues for future work.

2 Data

Geo. P. Rowell and Co.'s American Newspaper Directory was a directory to US newspapers and periodicals published in the mid-to-late 19th century, first published in 1869. 14 total[2] scanned editions were used to assemble a dataset of American newspapers over time from 1869 to 1890.[3] This contains newspapers listed in *Rowell's*, with their location, circulation, frequency of distribution, and most importantly, editors and owners for each year in the dataset.

Figure 1: Example of an entry in *Geo. P. Rowell and Co.'s American Newspaper Directory*


ABBEVILLE, C. H., *Henry Co.*, 500† pop., 90 m. S. E. of Montgomery, and 15 W. of Fort Gaines, Georgia.

HENRY CO. REGISTER; Fridays; democratic; four pages; size 22x32; subscription $2; established 1866; James W. Oates, editor; S. A. Fackler & Co., publishers; circulation 480, estimated.

The actual content of these newspapers was sourced from *American Stories* (Dell et al.,

[2] Not all editions in the years from 1869-1890 were available for download. The following editions were used: 1869, '71, '72, '73, '76, '77, '78, '79, '80, '82, '83, '84, '85, '90.

[3] See Appendix A for a description of how text was extracted from the scans, and associated difficulties/limitations of the data.

2023). The *American Stories* dataset is a collection of full article texts extracted from historical U.S. newspaper images. It includes nearly 20 million scans from the public domain Chronicling America collection maintained by the Library of Congress. Every scan from each year between 1869-1890 was included in the dataset.

To quantify topic coverage, I count the total number of appearances of a subset of keywords in headlines for a given paper in a given year (e.g., "strike" corresponds to the "labor_workers" topic). The keyword appearances are then multiplied by 1000 and divided by the total number of tokens to find the total number of keyword appearances per thousand tokens. This number is then scaled by the OCR grade[4] of a given paper in a given year, to attempt to account for the variable quality of the scanning and OCR text extraction process, i.e., for the possibility of a lower count of keyword appearances being attributable to higher percentages of garbled text rather than true differences in coverage.

Because calculating a baseline for a paper's "identity" (i.e., it's historical coverage patterns) relies on averaging multiple years of data, only newspapers with at least 3 years of pre-treatment data were considered, which reduces the total number of papers included in this study to 156.

I determine the timing of publisher turnover using Rowell's American Newspaper Directory. Since annual directories reflect information collected in the months prior to publication, a change in publisher listed in year t typically indicates a transition that occurred in year $t-1$. To correct for this reporting lag, I define the effective treatment year as the year prior to the first reported change ($t-1$). My results are robust to using the reporting year (t), though estimates of the onset of structural drift are naturally delayed by one period in the uncorrected specification.

[4]This OCR grade is calculated by counting the total number of appearances of a subset of common English words (mostly prepositions and articles), dividing this count by the total number of tokens overall to find a "rate of common word appearance", and then dividing this rate by a "baseline" rate of how often these words typically appear in the English language in any given text. This gives an approximate legibility grade to the digital scans of every paper in each year. It is of course possible that even within individual years, the legibility of scans of a given paper are highly variable, but it seems unlikely that this would result in systematic error.

3 Methodology

To estimate the impact of ownership changes on newspaper content, I employ a difference-in-differences design that quantifies the extent to which a newspaper's editorial focus diverges from its historical identity.

3.1 Measuring Structural Drift

The primary outcome variable is *Structural Drift*, a scalar metric designed to capture the magnitude of editorial repositioning regardless of the direction. I treat the topic composition of newspaper $i's$ coverage in year t as a vector $\mathbf{h}_{it} = (h_{it}^1, h_{it}^2, \ldots, h_{it}^K)$, where h_{it}^k represents the frequency of topic k (per 1,000 headlines).

First I establish a *baseline identity vector*, $\bar{\mathbf{h}}_i$, for each newspaper by averaging its topic frequencies during the pre-treatment period. Next, I calculate the structural drift, Y_{it}, as the Euclidean distance between the newspaper's content in year t and its baseline identity:

$$Y_{it} = \sqrt{\sum_{k=1}^{K} \left(h_{it}^k - \bar{h}_i^k\right)^2} \tag{1}$$

A higher value of Y_{it} indicates a greater deviation from the newspaper's historical norms. The advantage of this metric is that it is direction-agnostic; it captures editorial shifts whether a paper pivots toward commercial news, political partisanship, or sensationalism.

3.2 Identification Strategy

I estimate the causal effect of a change in publisher using a staggered difference-in-differences framework. I rely on *Rowell's American Newspaper Directory* to identify changes in publisher or editor. Because trade directories were typically compiled in the months prior to publication, a change listed in year τ likely reflects a transition that occurred in year $\tau - 1$. To correct for this reporting lag, I define the *Effective Treatment*

Year as one year prior to the first reported change.

I estimate the average treatment effect on the treated (ATT) using the following two-way fixed effects specification:

$$Y_{it} = \alpha_i + \delta_t + \beta \text{Post}_{it} + \varepsilon_{it} \tag{2}$$

where α_i represent newspaper fixed effects (controlling for paper specific or location specific trends) and δ_t represent year fixed effects (controlling for secular trends in the media market). The variable of interest, Post_{it}, is an indicator equal to one if year t is greater than or equal to the effective treatment year (i.e., $t \geq \tau_i - 1$). The coefficient β captures the average increase in structural drift attributable to the new publisher, relative to the natural evolution of the control group.

A potential threat to this identification strategy is the endogeneity of ownership turnover. Publisher changes are likely not usually exogenous: newspapers may be sold due to underlying financial distress, declining circulation, or management failures (although in some circumstances, such as some retirements or deaths, the change may be quasi-random). If these underlying conditions were already driving changes in editorial content prior to the sale (e.g., a failing paper attempting to pivot its coverage in a desperate bid for readers), our estimates might not capture the treatment effect of the new publisher. Thus, my dynamic event study explicitly tests for pre-treatment trends.

3.3 Dynamic Event Study

To test the validity of the parallel trends assumption and to examine the temporal evolution of editorial changes, I estimate a dynamic event study specification:

$$Y_{it} = \alpha_i + \delta_t + \sum_{j=-A}^{B} \gamma_j D_{it}^j + \varepsilon_{it} \tag{3}$$

Here, D_{it}^j is a set of dummy variables indicating that year t is j years relative to the effective treatment year ($\tau_i - 1$). The period $j = -1$ (one year prior to the effective transition) is omitted as the reference category to avoid multicollinearity.

The coefficients γ_j for $j < 0$ serve as a placebo test. This dynamic specification follows the event study framework developed in recent econometrics work (Sun and Abraham, 2021). Statistically insignificant estimates in the pre-period would support the parallel trends assumption, indicating that treated newspapers were not already undergoing structural changes prior to the ownership transfer. The coefficients for $j \geq 0$ capture the dynamic causal effect of the new publisher. I expect these coefficients to grow over time, reflecting a cumulative divergence from the old editorial regime.

Standard errors are clustered at the newspaper level to account for serial correlation within editorial tenures (Bertrand, Duflo, and Mullainathan, 2004).

4 Results

Table 1: Static Difference-in-Differences: Average Treatment Effect on Structural Drift

	Dependent Variable: Structural Drift
Post_{it}	10.4219^{***}
	(1.9575)
	[6.5851, 14.2586]
Newspaper Fixed Effects	Yes
Time Period Fixed Effects	Yes
R-squared	0.5154
Observations	2449

Notes: This table presents difference-in-differences estimates of the average treatment effect on structural drift. Standard errors are clustered at the newspaper level and reported in parentheses; 95% confidence intervals in brackets. Statistical significance at the 90%, 95%, and 99% confidence level denoted "*", "**", and "***".

Table 2: Event Study: Dynamic Effects on Structural Drift

	Dependent Variable: Structural Drift			
Event Time (k)	Coefficient	Std. Error	95% CI	P-value
$k = -5$	-0.1098		$[-3.8509, 3.6312]$	0.9541
$k = -4$	-1.3259		$[-4.4465, 1.7946]$	0.4050
$k = -3$	-2.6132^{*}		$[-5.5691, 0.3428]$	0.0832
$k = -2$	-3.1517^{**}		$[-5.9635, -0.3398]$	0.0280
$k = -1$		(Reference Period)		
$k = 0$	9.3294^{***}		$[3.7796, 14.8793]$	0.0010
$k = 1$	7.1247^{***}		$[3.3358, 10.9137]$	0.0002
$k = 2$	7.9342^{***}		$[4.2015, 11.6670]$	0.0000
$k = 3$	10.8576^{***}		$[6.3386, 15.3765]$	0.0000
$k = 4$	10.9673^{***}		$[5.9449, 15.9898]$	0.0000
$k = 5$	10.3075^{***}		$[5.5510, 15.0640]$	0.0000
Newspaper Fixed Effects		Yes		
Time Period Fixed Effects		Yes		
Observations		2449		

Notes: This table presents event study estimates of the dynamic treatment effects on structural drift. The omitted category is $k = -1$ (one period before treatment). Standard errors are clustered at the newspaper level. Statistical significance at the 90%, 95%, and 99% confidence level denoted "*", "**", and "***".

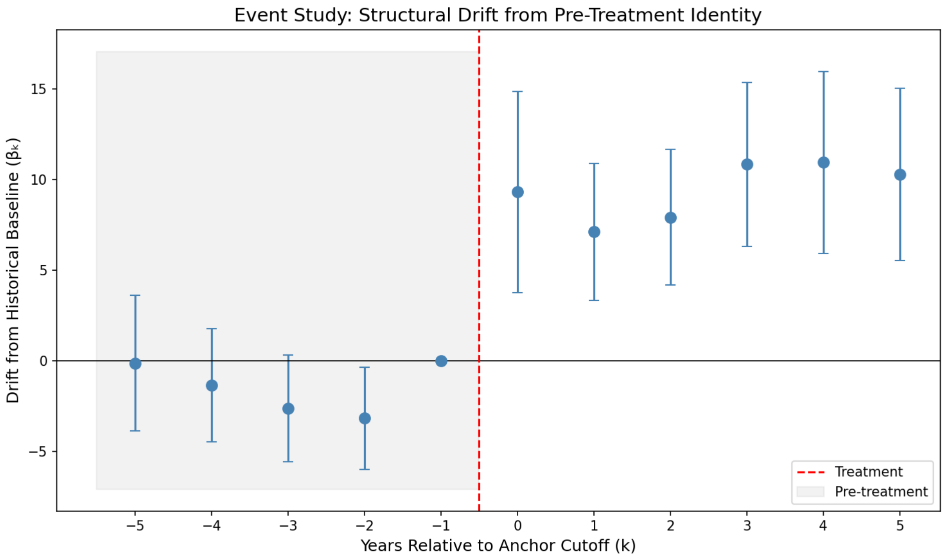


Figure 2: Event Study

Table 1 presents the regression estimates for the effect of publisher turnover on newspaper content. The results from the static difference-in-differences model are reported. The estimated coefficient for the variable of interest, β, is 10.42 (SE = 1.96). This estimate is statistically significant at the 1% level ($p < 0.001$) and indicates an average increase in structural drift of 10.42 units in the post-treatment period relative to the control group. The model has an R^2 of 0.515.

Table 2 reports the coefficients from the dynamic event study, with $k = -1$ serving as the omitted reference category. In the pre-treatment period, the coefficients for $k = -5$ ($\gamma_{-5} = -0.11$) and $k = -4$ ($\gamma_{-4} = -1.33$) are small in magnitude($p > 0.10$). In the effective treatment year ($k = 0$), the coefficient rises to 9.33 and is statistically significant at the 1% level. In the subsequent post-treatment years, the coefficients remain positive and significant, increasing from 7.12 at $k = 1$ to 10.86 at $k = 3$, before stabilizing at approximately 10.31 by $k = 5$.

The negative coefficients in the pre-treatment period ($k = -3, -2$) indicate that, prior to the transition, treated newspapers were actually more consistent with their historical baselines than the control group. This suggests that the ownership change did not follow a period of identity crisis or erratic coverage; rather, these papers were exceptionally stable before the transition. The abrupt reversal at $k = 0$, moving from negative pre-trends to a large, positive, and sustained drift, reinforces the conclusion that the new publisher was the primary catalyst for the structural break in editorial focus.

5 Conclusion

This essay examined the relationship between ownership turnover and news coverage in American newspapers from 1869 to 1890. Using a novel dataset that pairs Rowell's American Newspaper Directory with the American Stories digital archive, I quantified editorial shifts through a "structural drift" metric based on Euclidean distance. The results demonstrate that a change in publisher or editor serves as a significant catalyst for content reorganization, resulting in an average post-treatment drift of 10.42 units.

Notably, the event study reveals that treated newspapers were significantly more stable and consistent with their historical identities in the years immediately preceding a sale compared to the control group. This suggests that ownership changes in this era did not typically occur at papers already in the midst of identity crises or erratic content shifts. Instead, the arrival of a new publisher initiated a sharp and sustained departure from a previously stable baseline.

These findings suggest that while the 19th-century press was increasingly driven by advertising revenue and consumer demand, the specific identity of the owner remained a powerful determinant of topical focus. This research provides a framework for further investigating the "agency" of media owners during the Gilded Age and offers a methodology for capturing structural media shifts that are distinct from traditional measures of ideological slant.

6 References

Baldasty, Gerald J., 1992: *The Commercialization of News in the Nineteenth Century* (Madison: University of Wisconsin Press).

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan, 2004: "How Much Should We Trust Differences-in-Differences Estimates?", *The Quarterly Journal of Economics*, 119(1), 249-275.

Dell, Melissa, Jacob Carlson, Tom Bryan, Emily Silcock, Abhishek Arora, Zejiang Shen, Luca D'Amico-Wong, Quang Duy Le, Pablo Querubin, and Leander Heldring, 2023: "American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers", *Advances in Neural Information Processing Systems*, 36.

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

Leung, Tin Cheuk, and Koleman Strumpf, 2024: "Disentangling Demand and Supply of Media Bias: The Case of Newspaper Homepages", *CESifo Working Paper No. 10890.*.

Petrova, Maria, 2010: "Newspapers and Parties: How Advertising Revenues Created an Independent Press.", *The American Political Science Review*, 78(1), 35-71.

Sun, Liyang, and Sarah Abraham, 2021: "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects", *Journal of Econometrics*, 225(2), 175-199.