A reproduction of Newspapers in Times of Low Advertising Revenues*

Kar Hian Ong

22 December 2020

Abstract

In a rapid change industry, the newspaper industry constantly changes their ways to appeal to the readers. This paper uses observational data to investigate the effects of advertising revenue on French newspapers. We found that there is a lot of difference in the targeted variables for national newspapers compared to local newspaper. This paper has implication for the newspaper industries and how we look at this industry in the future.

Keywords: Newspapers, Ordinary least squares, Advertising, Journalist, difference in difference

I use R Core Team (2020), Ushey et al. (2020), Wickham et al. (2019), Wickham and Miller (2020), Hugh-Jones (2020) and Wickham and Seidel (2020) to complete this report.

1 Introduction

With the rise of the internet, the purchase of newspaper is in the decline. This is because not many people are willing to pay for things. With other forms of consuming news such as television and the internet, advertisers are more unwilling to pay newspapers to advertise their content. 1991 is when the internet went live to the world. Internet speed was not great then. At the turn of the century, internet speed went up and people's preferences from reading the news from the newspapers or television channels to reading the news from the internet. Advertisers have moved their ads online where they can get a bigger reach of people. Even traditional news outlet starts to move online as the way people consume news change.

In this paper I attempt to reproduce Newspapers in Times of Low Advertising Revenues by Charles Angelucci and Julia Cagé (Angelucci and Cagé (2019a)). This paper analyses the relationship between advertising revenue and newspapers' choices regarding the size of their newsroom, the quantity of news to produce and their pricing strategies. Their argument is based on the introduction of advertising on television which said to be affecting national newspapers than local newspapers. Hence, they perform a difference in difference analysis on the dataset. So, the local newspapers will be the control group while the national newspaper will be the comparison group.

Within this paper I attempt to use methods mention by the author of the paper.

The remaining section is structured as follows: section 2 will talk about retrieval and cleaning the data for the replication. section 3 will introduce the model, section 4 we will get to see the results of our modeling, section 5 will be discussion on the findings and weaknesses.

---

*Code and data are available at: https://github.com/karhian/s4-R-5

2 Data

The replication data is obtained from openicpsr (Angelucci and Cagé (2019b)). I ran through 01-data_cleaning-newspaper script to obtain the cleaned data.

The author of the paper collected data from various sources. On the data on the prices, circulation and revenue, those data are non-publicly available data obtained from the French Ministry of Information. The dataset includes 68 local newspapers and 12 national newspapers circulating between 1960 and 1974.

The number of journalists is obtained from the non-publicly available paper records of the e "Commission de la carte d'identité des journalistes professionnels" (CCIJP). The dataset includes information about journalists for 63 out of the 68 local newspaper and information about journalists in 11 out of 12 national newspapers. The number of newspaper journalists is one of the variables used as a proxy for the newspapers' quality. Besides data about newspaper journalist, data about television journalists is also collected from the French television and radio broadcasting agency (ORTF). The number of television journalist will be used as a variable as proxy for television quality. The data only account for journalist of newspapers and television that holds press card.

The data for advertising prices is the official list price per column inch of advertising space provided by "Tarif Media" an annual publication that provides information regarding advertising space. The author of the paper uses the price for the front-page ads as it is the place where it's frequently observed. Next, to get the quantity of advertising per issue, the data is collected from the paper version of the newspapers available in the French National Library.

Data about newshole comes from the content data collected by the author. The content is also being separated by hard story news and soft story news.

Lastly, the readership data on the other hand comes from the Centre d'Etude des Supports de Publicité (CESP). It is an association composed of main companies active in the advertising industry. The source of readership data comes from a survey conducted by CESP. The survey methodology for CESP is as follows, according to the online appendix, the target population is French citizens aged 18 or more living in metropolitan France. The sample size is between 250,000 and 300,000 individuals depending on year. The survey's main objective is to collect information about readership habits of French citizens at the time of the survey conducted using a questionnaire.

With the information about how the data is obtained by the author of the paper, I made an attempt to plot the data.
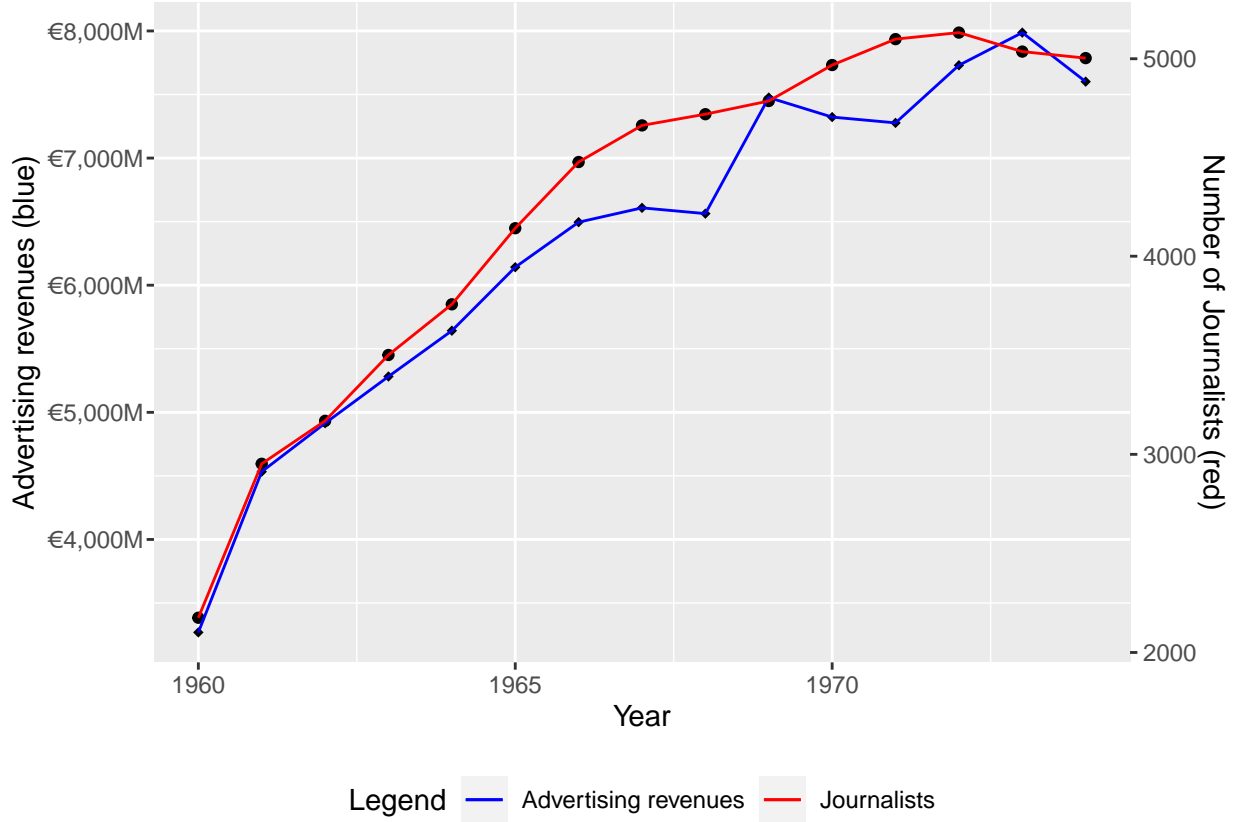
Figure 1: Newspaper Advertising Revenues (in euros) and Number of Journalists for french newspapers from 1960 to 1974

Figure 1 is a replication for figure 1 in the paper but I used the French newspaper data instead to do a comparison. It follows a similar trend where increase in advertising revenue increases the number of Journalists.

3 Model

The model given by the author is

$$y_{n,t} = \alpha + \beta_1(D_{\text{after}} * D_{\text{national news}}) + \lambda_n + \gamma_t + \epsilon_{n,t} \tag{1}$$

where $y_{n,t}$ is the outcome of interest, D_{after} is a switch where it switches on for observation after 1967, D_national news) is an indicator variable for national news, λ_n is the fixed effects for newspaper, γ_t is the time dummies and $\epsilon_{n,t}$ is newspaper year shock.

I will use the natural log of the value of the outcome of interest instead while modeling the dataset. This is so that the outcome value will be smaller. The equation is as follows:

$$\ln(y_{n,t}) = \alpha + \beta_1(D_{\text{after}} * D_{\text{national news}}) + \lambda_n + \gamma_t + \epsilon_{n,t} \tag{2}$$

We want to investigate the different models to look at effects on the advertising side, reader side and the quality of the newspapers based on equation (2).

4 Results

We start off by analyzing the advertisement side of the newspaper industry in France.

Table 1: Advertising Side

	Advertising revenues	Ad revenues div circulation	Ad price	Ad space
after_national	-0.23 ***	-0.15 ***	-0.31 ***	0.01
	(0.03)	(0.03)	(0.07)	(0.05)
year	0.05 ***	0.04 ***	0.04 ***	0.02 ***
	(0.00)	(0.00)	(0.00)	(0.00)
N	1052	1048	809	1046
R2	0.99	0.90	0.89	0.72
logLik	345.34	449.52	-277.71	-164.01
AIC	-526.68	-735.05	705.43	478.02

*** p < 0.001; ** p < 0.01; * p < 0.05.

We focus on the red text in table 1. We found that there is a decrease in advertising revenue, total advertising revenue normalised by circulation and ad price in national newspaper compared to local newspapers. We have found no statistical significance for advertising space.

Next lets take a look at the reader side of the dataset.

Table 2: Reader side

	Subscription price	Unit price	Circulation	Share of sub	Revenue from sales
after_national	-0.04 *	0.06 **	-0.06 **	0.19 ***	-0.06 *
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
year	0.05 ***	0.05 ***	0.01 ***	-0.01 ***	0.05 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	1044	1063	1070	1072	1046
R2	0.88	0.87	0.99	0.97	0.99
logLik	882.14	907.28	759.57	321.91	451.11
AIC	-1600.28	-1650.57	-1355.15	-477.81	-738.22

*** p < 0.001; ** p < 0.01; * p < 0.05.

For the red text in table 2, the subscription price, circulation and revenue decreased for national compared to local. while unit price and share of sub increased.

5 Discussion

5.1 Weakness

Since all the data used by the author of the paper is by external sources, the data may be bias to the data collector such as data favoring the data collector especially survey conducted by the CESP. It is stated that the CESP is association composed of main companies active in the advertising industry, so the survey reports may be favouring advertising companies. Since the paper I am reproducing is looking at the effects on advertising on the quality of newspapers there may be bias on the results.

Besides that, I could not find any information about the way CESP handles non-responses. The authors mention that they digitalise the report for the years 1957, 1962, 1967, 1968, 1969, 1970, 1972, 1974 as the paper was in paper format. They did not provide the digital copy in the replication data hence I am unable to get the details further from what is provided in the online appendix.

The way the author collects the data is not reproducible. They are either non-publicly available records or only available in paper copy. This makes it hard to verify the genuineness of the data.

References

Angelucci, Charles, and Julia Cagé. 2019a. "Newspapers in Times of Low Advertising Revenues." American Economic Journal: Microeconomics 11 (3): 319–64. https://doi.org/10.1257/mic.20170306.

———. 2019b. Replication Data for: Newspapers in Times of Low Advertising Revenues. Nashville, Tn: American Economic Association [Publisher], 2019. Ann Arbor, Mi: Inter-University Consortium for Political and Social Research [Distributor], 2019-12-07. https://doi.org/10.3886/E116438V1.

Hugh-Jones, David. 2020. Huxtable: Easily Create and Style Tables for Latex, Html and Other Formats. https://CRAN.R-project.org/package=huxtable.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ushey, Kevin, JJ Allaire, Hadley Wickham, and Gary Ritchie. 2020. Rstudioapi: Safely Access the Rstudio Api. https://CRAN.R-project.org/package=rstudioapi.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." Journal of Open Source Software 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2020. Haven: Import and Export 'Spss', 'Stata' and 'Sas' Files. https://CRAN.R-project.org/package=haven.

Wickham, Hadley, and Dana Seidel. 2020. Scales: Scale Functions for Visualization. https://CRAN.R-project.org/package=scales.