

Replication of the Revenue Difference of Differences Analysis in ‘Newspapers in Times of Low Advertising Revenues’, by Charles Angelucci and Julia Cagé

Wenxi Li

07/12/2020

Abstract

This paper summarises some of the main findings of ‘Newspapers in Times of Low Advertising Revenues’, by Charles Angelucci and Julia Cagé. In the study, we found that there’s a negative shock on the advertising side, the reader side, and the quality of newspapers. In addition, we also find that the composition of the readership changes after the arise of TV advertisements. The findings of the paper could serve as a reference to help the newspaper industries or perhaps other media outlets making decisions when a threat to the industry arises.

Keywords

difference-in-differences study, newspaper industry, advertisement, replication of study, linear regression model

Introduction

In this paper, we are going to replicate some of the main findings of the study ‘Newspapers in Times of Low Advertising Revenues’, by Charles Angelucci and Julia Cagé.

Newspaper industries have been experiencing a hard time in recent years; their advertising revenue has been decreasing over the years. In the study, the authors are trying to investigate the effect that decreasing advertising revenue will bring to the newspapers.

However, there are clearly many factors that could influence the advertising revenue, which could potentially affect the result. In the study, the authors find out a brilliant way to overcome this problem, that is, using a data set containing a number of different French newspapers’ data from 1960 to 1974. Notice that in 1968, there was an introduction of TV advertisement, which could cause a negative effect on the newspaper’s advertising revenue. Specifically, the author found out that local newspapers’ advertising revenues are much less affected by the TV advertisement than the national newspapers. Therefore, the study uses a difference-in-differences analysis, where we consider the local newspaper as the control group and the national newspaper as the treatment group.

In the following study carried out, we are going to investigate and predict the effect of the TV advertisements bring to the advertising-side of the newspaper, reader-side of the newspaper, quality of newspapers, and the composition of the readership. Specifically, for the advertising-side, we are going to investigate how the advertising revenue and quantity of advertisement change, for the reader-side, we investigate how the pricing strategy, as well as the number of subscriptions change.

For the rest of the paper, we will include a methodology part, where we will describe the data set that is used in the study as well as the chosen model. For the main findings, we are going to discuss in the results part. We will also include a summary as well as some next steps and weakness in the discussion section.

Data

The original study uses a number of different data sets throughout the whole paper, but its main analysis only uses one major dataset. For our replication of the original study, we are going to use exactly the same one.

The major dataset used is called ‘Angelucci_Cage_AEJMicro_dataset’. The dataset contains numerous variables regards different aspects of the newspaper companies in France from 1960 to 1974. The dataset contains 50 variables regarding a different aspect of the newspapers, where one of them was calculated and constructed by ourselves later. Notice that different variables were collected from different sources.

Data on prices, revenue, and circulation were collected from the French Ministry of Information’s non-publicly available records in the National archives. A sample of 68 local newspapers is included, which was a fairly large proportion of the total number of local ones. Meanwhile, all the 12 national newspapers are included.¹

The number of journalists was collected from “Commission de la carte d’identité des journalistes professionnels” (CCIJP), the organization that issues press cards to journalists in France since 1936. This time we have the sample data on 63 out of 68 local newspapers and 11 out of 12 of the national nespapers.²

Advertising price and quantity data were first collected from the official list price per column inch of advertising space, but it was realized by the authors that the discount on advertising price is common. So instead, they calculated the advertising price per unit quantity directly using the data from the paper version of the newspapers available in the French National Library.

They directly studied the content of the newspaper issued in two complete weeks per year and calculated the amount of space that the news occupies. The result is collected in the dataset named Newswwhole.³

As mentioned above, a new variable `ra_cst_div_qtotal`(average advertising revenue per circulation) is constructed, and it’s calculated by the `ra_cst`(total advertising revenue) divided by `qtotal` (total circulation).

The graph below summarises the total advertising revenue of local and national newspapers before and after 1967.⁴ Clearly, we see that local newspaper are not influenced much by the introduction of TV advertisement, while the national newspaper experienced a slight decrease in the advertising revenue.

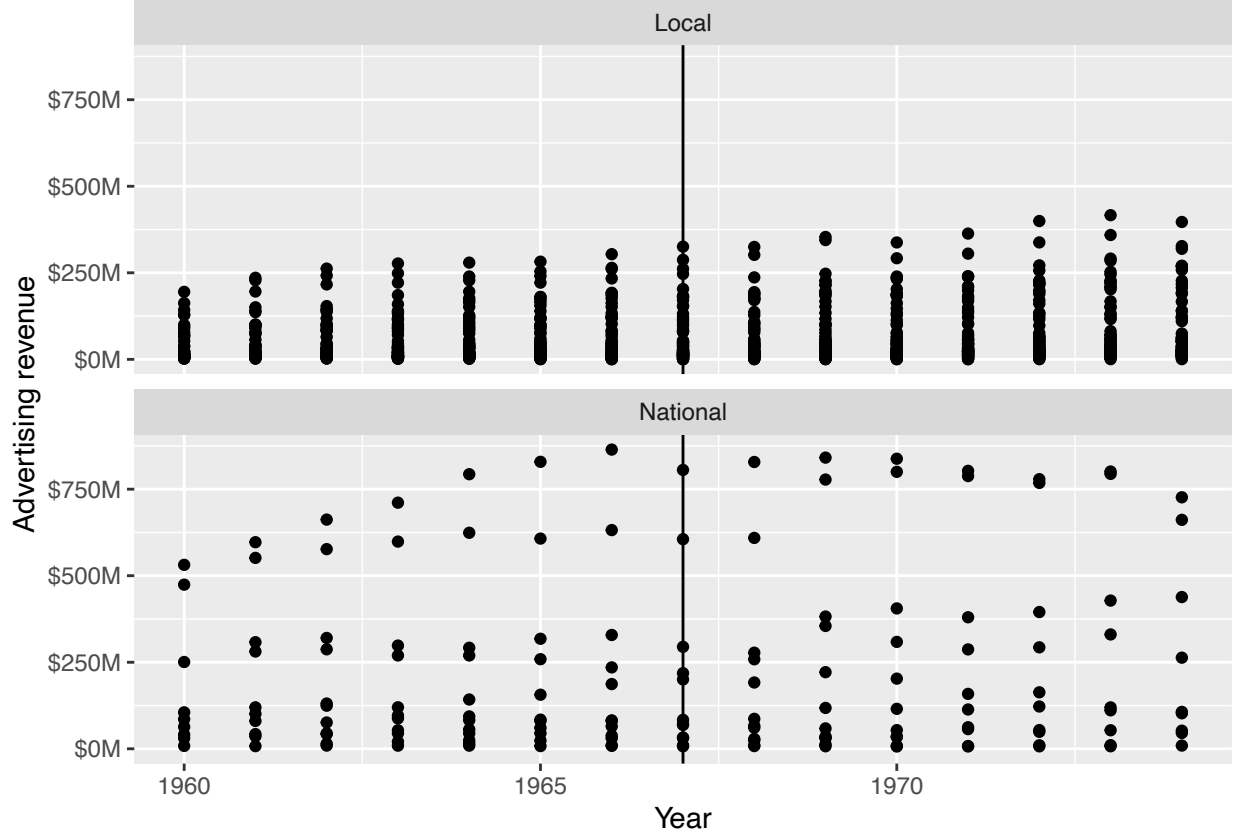
In a nutshell, this dataset is brilliant to carry out the analysis; it has numerous variables that we can look into, and the sample size is large enough to carry out a sound analysis. In addition, we have an equal spread of data about the year 1967 in which the treatment(TV ads) came into place, which is in favor of the study of the effect brought by the TV advertisements.

¹This is already a fairly large and convincing sample, so omitting the few local newspapers is not going to influence the result too much.

²Again this a pretty convincing sample population size.

³This time, an extra of 37 local newspapers were studied as well as all the national newspapers. The target population and sampled population are about the same.

⁴Code for graph credit to Rohan Alexander from https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html#results



Model

In this paper, we are going to use R studio to carry out the following analysis.

The model that is used to predict the effect of television advertisements is a simple linear regression shown below.

A linear regression model is most appropriate here since we expect the response variable to have a linear relationship with the predictor variables.

$$\log(y_{n,t}) = \alpha + \beta_1(D_{after} \times D_{nationalnews}) + \lambda_n + \gamma_t + \epsilon_{n,t},$$

Notice that on the left-hand side, we are using a transformation with a log. This is because that the main predictor ($D_{after} \times D_{nationalnews}$) has only two values, 0 and 1, and it is not likely that it will give us a linear relationship. So we transform y so that it will look more like a linear relationship.

So in the model $D_{after} \times D_{nationalnews}$ is a categorical variable. It equals one if the year of that data recorded is after 1967 and it's a national newspaper, and it equals zero otherwise.

γ_t represents the fixed year effect. So instead of being a time variable, we are going to change the time data to a list of integers.

Similarly, λ_n represents the fixed effect of newspapers, and we consider it as a categorical variable.

α simply represents the intercept, but it doesn't have any practical meaning here since we don't expect the fixed effect for year and newspapers to be zero.

$\epsilon_{n,t}$ could be simply treated as the standard errors here.

Meanwhile, our main interest of focus is β_1 , as it directly predicted how much $\log(y_{n,t})$ ⁵ changes due to the arrival of the TV advertisement.

⁵Here the response variable could vary between reader side variables, advertising side variables, quality variables and

In the original study, the composition of readership is being analyzed in two ways, education, and occupation. In our replication of the paper, we are only going to investigate the occupation composition, as it is a better indication of the individuals' news taste as well as their disposable income.

These two fixed effects are essential in our analysis since it prevents variations due to time and different newspapers. Therefore, with the help of the fixed effects, we can measure the annual effects that TV advertisements would bring to the newspaper industry.

As shown in the data part, the total advertising revenue received by local newspapers is much less influenced by TV ads than the national newspapers. So we are assuming that the trend of all the variables of both national and local newspapers would be the same before 1967, that is, in the absence of the treatment. After the treatment is introduced, we are essentially carrying out a difference-in-differences study where the local newspapers are the base group and the national newspapers are the treatment group.

Therefore, as mentioned above, we are using the predictor $D_{after} \times D_{nationalnews}$ as a dummy variable. All it does is setting the predictor of the local newspaper to be 0, and the national newspaper after 1967 to be 1. Then the result ($\log(y_{n,t})$) is predicted by the difference between the local newspaper group and the national newspaper group.

Notice that for $\log(y_{n,t})$, we could test numerous different variables in different aspects, but we are strictly using the dataset that the original paper used. When we consider the sources that the authors got their dataset, it would be extremely time-consuming to obtain another set of variables.

Results

We summarise the main results we got using the model above in the four tables below.

First, we look at the effect of the advertising side (table1). We find that the advertising revenue per circulation decreases by 15 percent compared to the local newspapers, and this directly causes the total advertising revenue to fall by 23 percent. When we only look at the listed advertising price, we see that it decreases the most by 31 percent. As for advertising space, we do not find it changes statistically significant.

Advertising Side

	Advertising Revenue	Advertising Revenue Circulation	Advertising Price	Advertising Space
Estimate	-0.23	-0.15	-0.31	0.02
Standard Error	0.03	0.03	0.07	0.05
Newspaper Fixed Effect	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes
P-value	<0.05	< 0.05	< 0.05	0.78

table 1

Next, we take a look at the reader side variables (table2). We see that the changes in all the variables are statistically significant. Interestingly, we see that the subscription price decreases while the unit price increases when compared to the local newspapers. We also see that the circulation, as well as the revenues from sales, decrease while the share of subscribers increases by 19 percent.

Reader Side

	Subscription price	Unit price	Circulation	Share of subscribers	Revenues from sales
Estimate	-0.04	0.06	-0.06	0.19	-0.06
Standard Error	0.02	0.02	0.02	0.03	0.03
Newspaper Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
P-value	<0.05	< 0.05	< 0.05	<0.05	<0.05

table 2

Then we look at the effect on the quality of the newspapers (table 3). Out of all the five variables we analyzed, we only find the number of journalists and the number of page changes statistically significantly. We see that the number of journalists decreases by 20 percent and the number of pages decreases by 2 percent. However, as mentioned before, the authors argue that there are potentially many other variables that could describe the quality of a newspaper. What is included in this study are limited.

Quality

	Number of journalists	Average payroll	Number of pages	Newshole	Share of hard news on front page (percent)
Estimate	-0.20	0.06	-0.02	-0.04	-0.06
Standard Error	0.03	0.04	0.02	0.02	0.04
Newspaper Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
P-value	<0.05	0.2	< 0.05	0.07	0.19

table 3

Lastly, we look at the effect the TV advertisements bring to the composition of the readership's occupation (table4). we could clearly see that the national newspapers are now having more readers with less disposable income compared with the base group-the local newspapers. Specifically, we see that the percentage of employees reduces by astonishingly 8.73 percent, and the percentage of laborers increases by 4.83 percent.

Readership

	Farmers	Artisans_and_shopkeepers	Senior executives	Employees	Laborers
Estimate	2.50	-0.11	-1.01	-8.73	4.83
Standard Error	0.49	0.33	0.37	0.61	0.04

	Farmers	Artisans_and_shopkeepers	Senior executives	Employees	Laborers
Newspaper Fixed Effect	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes
P-value	<0.05	0.74	< 0.05	<0.05	<0.05

table 4

Discussion

Overall, the results we have are very similar to what the original study finds; from 1967 onwards, we certainly find many variables are influenced by the arrival of the TV advertisements.

Notice that in the replication of the study, we are using a benchmark of 0.05, instead of 0.10 which the original paper uses. Therefore, the final interpretation of the results may slightly differ from the original paper.

Interpretation of some of the results

When we look at table1, there is a difference between the change in list price and the actual advertising price per circulation; the list price is much higher. This is caused by the discount effect that we mentioned before. This is another indication that the use of price per circulation in the model is more preferable since the list price is not finalized which has the potential to change over time.

Table 2 suggests that the subscription price decreases while the unit price actually increases. Charles Angelucci and Julia Cagé suggest that this is the evidence of increasing price discrimination posed by the national newspapers, which means they probably charge much higher to those who are loyal customers of that newspaper.

Robustness of the model

In the original paper, the authors prove that the model is robust to many potential threats. For example, all the national newspapers and local newspapers are oligopolies⁶, and the model assumes they have equal degrees of competition, which is not very likely for this to happen in real life. However, the model is proved to be robust to the different degrees of competition.

Some of the weakness

One of the weaknesses of the study is that, when we use the difference-in-differences model, we are essentially assuming that the local newspapers are not influenced by the arrival of TV ads. Therefore, the final results of all those changes in different variables may vary based on the effect that the threat brings to the local newspaper.

The other weakness, as mentioned by the authors is that we could see that the cost of transportation of national newspapers was generally higher than that of a local newspaper since a large portion of the national newspaper were being transported by train, which was more expensive than postal services. Therefore, given this fact, we could see the national newspapers were affected in numerous way, and we certainly need to consider them to obtain a more convincing result.

Another weakness is that we do not have enough datasets for the national newspapers since at that time there're only about 10 of them. Our sample population is enough compared to the total population that we are interested in, but it's not enough to carry out an accurate regression model,i.e, the total population is not enough.

⁶Oligopoly means the majority of the newspaper industry are consisted by a couple of large newspaper companies.

External Validity

As suggested by the authors, you could argue that the dataset is dated, and the French Newspaper System could be different from others'. However, we could see that even today, the newspaper industry is still one of the major media outlets. Furthermore, newspapers are still relying heavily on advertisement revenue and the advertisement revenue is decreasing with time, which is exactly the situation we have with our datasets.

On the other hand, subsidies might be more prevalent in France than in other countries, but subsidies didn't come out until the end of the time that we include in our dataset (1960-1974).

Next steps

As explained, the original dataset was collected from different paper sources, and the authors manually digitalize them to carry out the following analysis. Notice that not all the newspapers are being recorded at that time and bias in terms of different prices may occur. So, we could further gather more information regarding the French newspaper industry at that time if it is time-worthy.

In terms of the replication of the original study, we could further use the software that the original study use to see we could carry out the exact same result.

Appendix

All the code used are included in the .rmd file, which could be found at <https://github.com/wethanl/STA304-FINAL-PROJECT>

Reference

1. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>.
2. Angelucci, Charles, and Julia Cagé, 2019, 'Newspapers in times of low advertising revenues', American Economic Journal: Microeconomics, <https://www.openicpsr.org/openicpsr/project/116438/version/V1/view>.
3. Angelucci, Charles, and Julia Cagé. 2019. "Newspapers in Times of Low Advertising Revenues." American Economic Journal: Microeconomics, 11 (3): 319-64. DOI: 10.1257/mic.20170306
4. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." Journal of Open Source Software, 4(43), 1686. doi: 10.21105/joss.01686.
5. Hadley Wickham, Evan Miller, RStudio <http://haven.tidyverse.org>, <https://github.com/tidyverse/haven>, <https://github.com/WizardMac/ReadStat>
6. Hadley Wickham, Dana Seidel, RStudio <https://scales.r-lib.org>, <https://github.com/r-lib/scales>
7. Alexander Rohan https://www.tellingstorieswithdata.com/06-03-matching_and_differences.html#results