

Reproducing Angelucci & Cage’s ‘Newspapers in Times of Low Advertising Revenues’

Yuyan Xu

22 December 2020

Abstract

The present paper aims to replicate the results from Angelucci & Cage’s paper, and provides more validity to the finding that alternate advertising platforms tend to decrease the newspaper producers’ incentives to advertise, which leads to a decrease in journalistic-intensive content in the newspapers. Our findings duplicated the results in Angelucci & Cage that a decrease in advertising revenues has effects on newspaper’s contents and focus on readership.

Keywords

Advertising revenues; French newspaper; Journalistic-intensive content; Readership; Pricing strategies

Github Repository

Code for this analysis is available at: <https://github.com/Christinax86-tech/Final>

Introduction

The advancement in communication technology today provides general readers with more ways to access information. People can easily fetch news from TV, smartphones, tablets, etc. It is of interest that how newspaper industry change their method of delivering news, and how their advertising revenues changes corresponding with the newspaper content and pricing. In Angelucci & Cagé’s “Newspapers in Times of Low Advertising Revenues”, they explored the shock brought to the advertising side of French newspaper industry with the introduction of television advertising. Also, they used empirical analysis to discover the relationship between advertising revenues and the incentive of newspaper producers put journalistic-intensive contents in their publications (Angelucci & Cagé, 2015). They found that there was a decline in newspaper producers’ willingness to pay for advertising for attracting readers’ attention with the introduction of alternative advertising platforms; also, a decrease in advertising avenue lead to a corresponding decrease in journalistic-intensive content in newspaper (Angelucci & Cage, 2015). The present paper aims to reproduce these results using simple linear model with ordinary least squares estimators. Based on the results obtained from models, our findings were that a decrease in advertising revenues has effects on newspaper’s contents and focus on readership, which were a decrease in journalist-intensive contents and switch of focus on less educated readership. These findings duplicated the findings in Angelucci & Cage’s paper.

Data

The data would be the data used in Angelucci & Cagé’s paper: the information about newspaper prices, revenues, and circulation from the French Ministry of Information’s non-publicly available records in the

National archives, which includes 68 local newspapers and 12 national newspaper; the number of journalist from the non-publicly available paper records of the “Commission de la carte d’identité des journalistes professionnels” (CCIJP) , which includes 63 out of the 68 local newspapers for which we have revenue data, and 11 out of the 12 national newspapers; information about advertising prices and quantity of advertising from “Tarif Media,” and French National Library respectively. However, since we tried to duplicate their results, we only chose the variables that were of interest. They included two aspects: newspaper production side and readership side. Potential response variables in newspaper production included total circulation, number of pages in publication, news-hole, number of journalists; potential response variables in readership included: subscription price per issue, unit buyer price, percentage of share of subscribers, reader’s level of education.

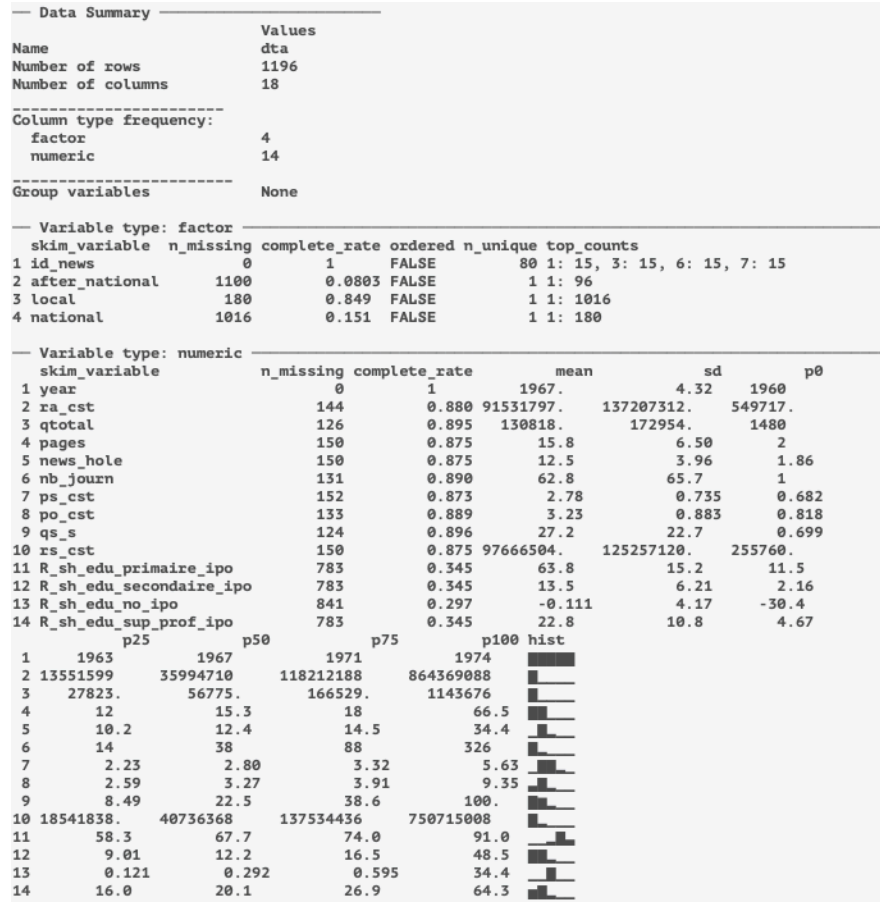


Figure 1 Overview of dataset

with variables of interest

From Figure 1, we took a overall view on the dataset contained with these variables of interest.

Model

The models we are interested in are based on the Ordinary Least Squares (OLS) method, which is a method commonly used to examine linear regression models :

$$\ln(Y_{n,t}) = \beta_0 + \beta_1 + \dots + \beta_n + \epsilon$$

where

- $\hat{\beta}_0$ is the OLS estimator of the intercept coefficient β_0
- $\hat{\beta}_1 \dots \hat{\beta}_n$ are the OLS estimators of the slope coefficients $\beta_1 \dots \beta_n$

- ϵ is a random error term

The rationale of the estimated models in present paper are based on the OLS simple regression function

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_n$$

By using the OLS estimation, we can minimize the sum of the squared residuals, in other words, minimize the squared prediction errors of the model. This means that we can obtain a more accurate model using OLS. In this case, Y represents the predictor variable, which is advertising avenue, X represents the response variables, which were selected based on the findings in Angelucci & Cagé, 2015. By doing this, we aimed to reproduce their results by examining the significance of the relationship between these response variables X and the predictor variable Y. We built two models, one is from the aspect of newspaper production, another is from the aspect of readership. Therefore, the response variables were selected respectively in terms of these two aspects, while the predictor variable, advertising avenue, stayed the same in each model.

First, we aimed to investigate the potential variables in newspaper production (total circulation, number of pages in publication, news-hole, number of journalists) that could be significantly influenced corresponding to the changes in advertising avenues.

```
##
## Call:
## lm(formula = ra_cst ~ qtotal + pages + news_hole + nb_journ,
##     data = dta)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -215025609  -24083481   1410071   19360215  416074583
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.412e+07  7.128e+06   1.980   0.048 *
## qtotal       2.126e+02  2.389e+01   8.896  <2e-16 ***
## pages       1.165e+07  6.425e+05  18.129  <2e-16 ***
## news_hole   -1.672e+07  1.056e+06 -15.840  <2e-16 ***
## nb_journ    1.204e+06  6.596e+04  18.262  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58600000 on 913 degrees of freedom
## (278 observations deleted due to missingness)
## Multiple R-squared:  0.8337, Adjusted R-squared:  0.8329
## F-statistic: 1144 on 4 and 913 DF, p-value: < 2.2e-16
```

Figure 2 The summary statistics for model1

We first included all the selected variables in model1 and the summary statistics indicated that they all have significant relationships with the predictor variables, since the p-values were extremely small. However, a best model should contain fewer number of variables if possible, since there is a chance that a variables had small p-value because of its co-effects with another variable contained in the model. In this case, if remove the other variable from the mode, as a result, the former variable would not have a significant relationship or its significant value decreases. Therefore, we removed variable of total pages of publications from the model in order to examine if there is a variable consequently has reduced significant level. Based on the summary statistics of model2, the p values basically remained unchanged. In the same way, we constructed the third model by removing variable, number of pages, from model1. The p-values of all the variables still remained unchanged. However, it is worth mentioning that no matter which variable we remove from model1, the value of adjusted R-squared decreases. This suggests that all the variables contained in model1 are significant to the fit of regression model, in other words, by adding each of the variable into the model,

the model became increasingly better. Therefore, for the newspaper production, variables which are total circulation, number of pages in publication, news-hole, number of journalists, can be affected by a change in advertising avenues.

However, the vif value of variable, number of journalists, was 5.18, which was higher than the threshold of 5. It means that this model had problem estimating the coefficients although its predictive ability was not affected. The vif values of model2 were good, which were all below the threshold of 5.

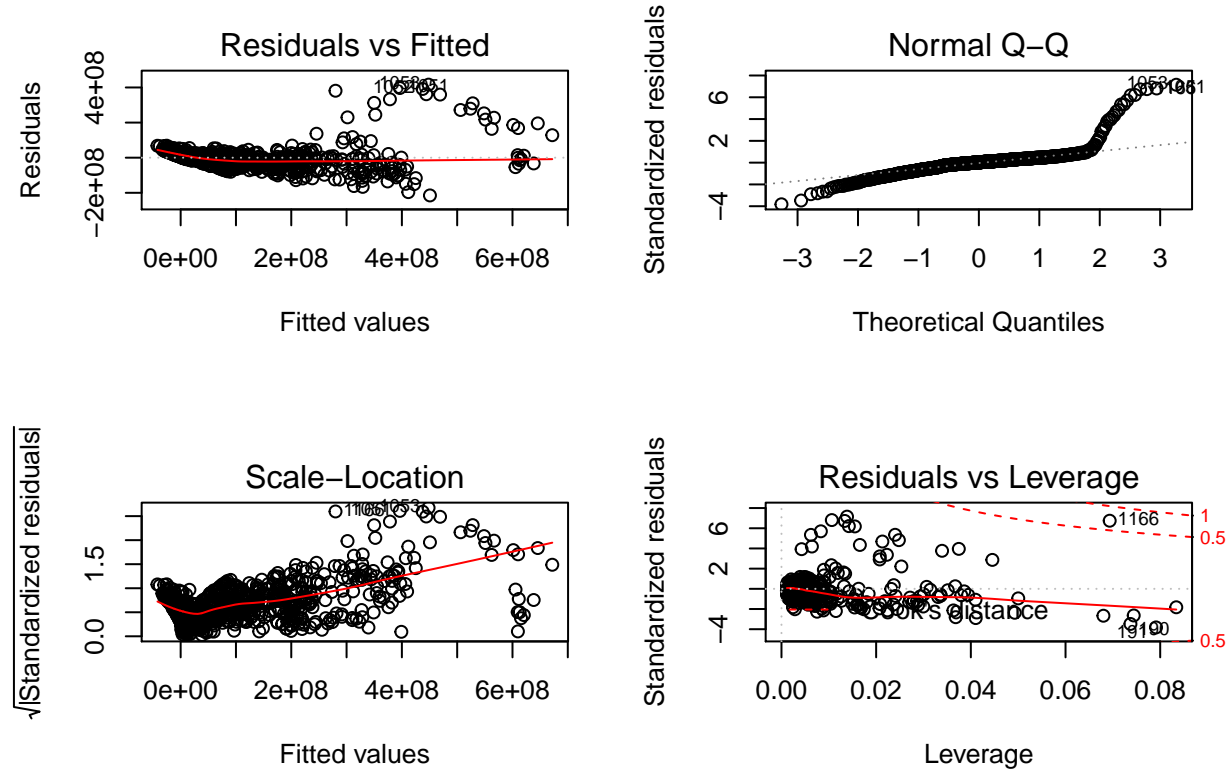


Figure 3 Plotting model2 and examining its validity

We also examined the validity of model2 by plotting the model in residuals vs. fitted graph, QQ plots, Scale-location plot and residual vs leverage plot. From the produced plots above, we can see that although there are some significant leverage points, the overall linear trend of model1 is clear.

Therefore, for newspaper production side, we decided to choose model2 as our estimated model, which contains variables: number of pages in publications, news-hole and number of journalists.

Second, we examined the potential variables in the reader side: subscription price per issue, unit buyer price, percentage of share of subscribers, reader's level of education. We removed the variables that showed no significant relationship with advertising revenues from model_1, and obtained model_2 with the rest of the variables: subscription price per issue, unit buyer price, percentage of share of subscribers, reader with primary education.

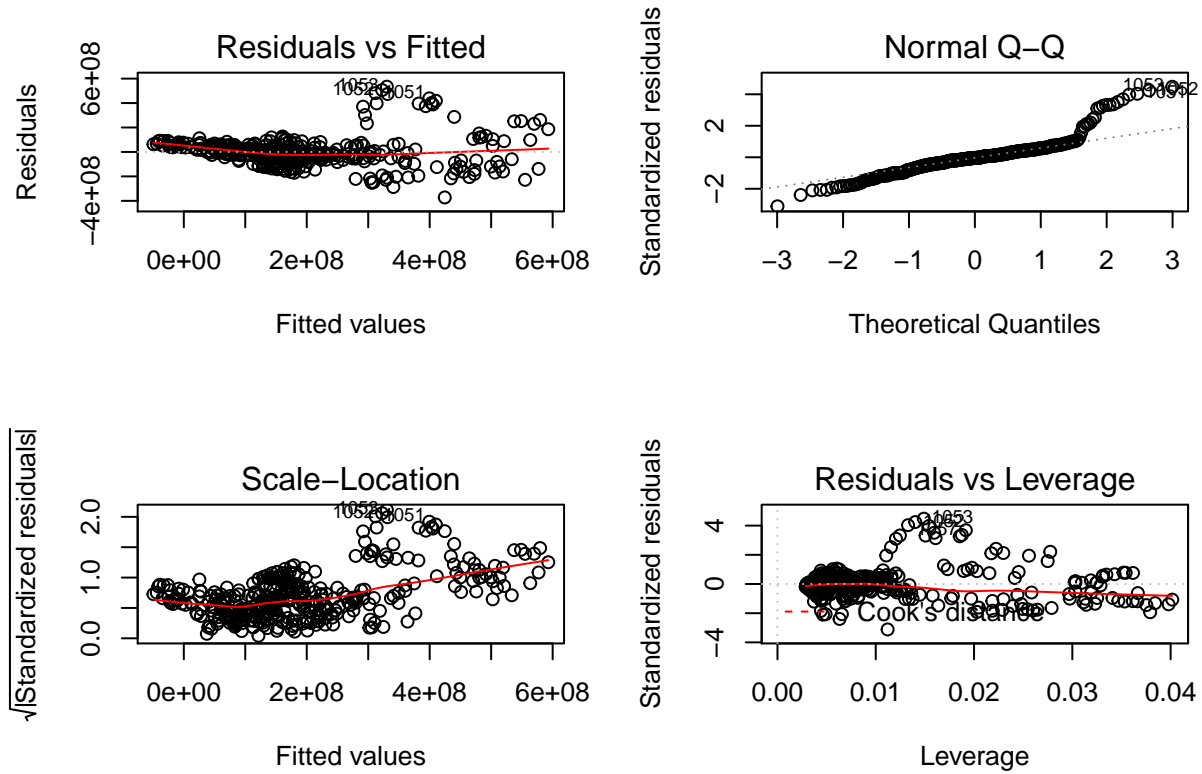


Figure 4 Plotting model_3 and examining its validity

Like what we did for model2 in newspaper production side, we also produced plots for model_3 in reader side. Also, the vif values for model_3 were all around 1, which means that the variables in model_3 did a good job on estimating coefficients.

Result

We have several key findings summarized from the model section. First, the number of pages in publications, number of news-holes, and number of journalists were all associated with advertising revenues for newspaper industries; Second, the percentage of subscribers, revenues from sale, reader's education level were all positively correlated with advertising revenues; Third, which is an interesting finding, according to the summary statistics of model1, the number of news hole had negative correlated relationship with advertising revenues.

Discussion

Our findings in present paper indicated several similarities with Angelucci & Cage's paper. First, we found that increase in advertising revenues would lead to subsequent increase in number of pages in publications and increase in number of journalists. This shows that increase in advertising revenues provides incentives for improving quality of the publications, as the number of journalists and newsrooms increase. Although our finding generally duplicated Angelucci & Cage's result that a decrease in advertising revenues would lead to a decrease in journalist-intensive contents in newspaper as the number of journalists decreases correspondingly (Angelucci & Cage, 2015), we found one difference with their findings: as the summary statistics of model1 indicated that the number of news hole had negative correlated relationship with advertising revenues. This means that as the advertising revenues increases, the number of news hole decreases.

	Number of journalists	Average payroll	Number of pages	Newshole	Share of hard news on front page (percent)
National \times Post-TV ad	-0.21 (0.06)	0.06 (0.06)	-0.03 (0.08)	-0.04 (0.08)	-0.11 (0.07)
Newspaper fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
R^2	0.50	0.28	0.61	0.52	0.11
Adjusted R^2	0.49	0.26	0.61	0.52	0.09
Observations	1,046	723	1,046	1,046	418

Figure 5 Findings in Angelucci & Cage's paper regarding advertising revenues and its effects on newspaper quality (Angelucci & Cage, 2015)

This finding was not corresponding to the suggestive evidence that provided in Angelucci & Cage's (See Figure 5): the number of news hole remained unchanged under the shock wave brought by alternative news platform, although the number of journalists went fewer, which indicated a switch from hard news to soft news (Angelucci & Cage, 2015). The reason that we had this minor difference in finding with Angelucci & Cage might be we did not include the consideration of time of year. If we analyze the changes after the year 1967 (which is the important year that television advertising started to compete with newspaper advertising).

	No diploma	Primary education	Secondary education	Professional education
<i>Panel A. Education</i>				
National \times Post-TV ad	0.66 (1.54)	1.55 (2.26)	-2.44 (0.95)	0.38 (1.40)
Newspaper fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
R^2	0.05	0.72	0.38	0.62
Adjusted R^2	0.02	0.71	0.35	0.60
Observations	413	413	413	413

	Farmers	Artisans and shopkeepers	Senior executives	Employees	Laborers	Inactive
<i>Panel B. Socio-professional category</i>						
National \times Post-TV Ad	2.70 (0.97)	-0.12 (0.62)	-1.40 (0.83)	-8.29 (2.15)	4.44 (1.37)	0.18 (1.15)
Newspaper fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.44	0.19	0.11	0.62	0.13	0.75
Adjusted R^2	0.42	0.16	0.07	0.60	0.10	0.74
Observations	413	413	413	413	413	413

Figure 6 Findings in Angelucci & Cage's paper regarding advertising revenues and its effects on readership (Angelucci & Cage, 2015)

Second, we found that newspaper tend to switch to focus on appealing to less educated reader group as the advertising revenues decreases. This result agreed with Angelucci & Cage's finding that the collapse in advertising revenues led to a switch of focus on readership followed by readjustment of content and prices (See Figure 6). However, it is worth mentioning that they separated the local and national newspaper and examined the effects on them separately. These findings were mostly related to national newspaper since

it seems that national newspaper was taken a huger blow compared to local newspaper (Angelucci & Cage, 2015).

Weakness and Next Step

There are several weaknesses in the present paper. We did not exactly duplicate Angelucci & Cage's analysis procedures. In their paper, they 1) compared the changes in advertising revenues before and after the introduction of alternative advertising platforms such as televisions, which is the year 1967; 2) considered the effects of decrease in advertising revenues on local and national newspaper separately. In the present paper, we only considered the decrease in advertising revenues on newspaper industry and readership, which was more general compared to Angelucci & Cage's paper. However, this also means that our findings were less sophisticated compared to Angelucci & Cage's paper since we failed to do the difference-to-difference analysis method as they did in their paper.

For future research, it is necessary to be more specific in terms of the research direction. It is recommended to use a different analysis method and see if the results duplicated that in Angelucci & Cage's paper.

Reference

- Angelucci, Cage. (2015). Newspapers in Times of Low Advertising Revenues. *American Economic Journal: Microeconomics*, 11(3). DOI: 10.1257/mic.20170306
- Luca Braglia (2016). RStata: A Bit of Glue Between R and Stata. R package version 1.1.1. <https://CRAN.R-project.org/package=RStata>
- Yihui Xie (2014) knitr: A Comprehensive Tool for Reproducible Research in R. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595
- Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- John Fox and Sanford Weisberg (2019). *An {R} Companion to Applied Regression*, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0
- Matt Dowle and Arun Srinivasan (2019). data.table: Extension of `data.frame`. R package version 1.12.8. <https://CRAN.R-project.org/package=data.table>
- Martin Elff (2020). memisc: Management of Survey Data and Presentation of Analysis Results. R package version 0.99.27.3. <https://CRAN.R-project.org/package=memisc>
- Simon Urbanek (2013). png: Read and write PNG images. R package version 0.1-7. <https://CRAN.R-project.org/package=png>