# Finding the impact of reduction of advertising revenues to newspapers

Huynh, Vu Minh Hoang

22th Dec, 2020

## Finding the impact on the reduction of advertising revenues to newspapers agency

**Huynh, Vu Minh Hoang**

**22th December, 2020**

Code and data supporting this analysis is available at: https://github.com/axelhvmh/final.git

## Abstract

The newspaper industry is in economic turmoil in recent years where there are huge competition with the Internet, competing for its potential readers as well as sources of revenue. We found that there is a positive relationship between the advertising revenue and the quality of content of a newspaper agency.

Keywords: Simple linear regression, advertising revenue, number of journalists

## Introduction

Since the mid 2000s, newspaper agencies have seen a drop in their revenues, as well as the number of journalists under their employment. A variety of reasons such as: the change in consumer preference or the rapid rise of the Internet, could contribute to the change in quality as well as revenues for a newspaper. With the shrinkage of the industry, there are growing concerns as to whether the quality of a newspaper has decreased or not because of that.

The paper "Newspapers in Times of Low Advertising Revenues" published by Charles Angelucci and Julia Cagé has extensively talked about the effects of how low advertising revenue affect a newspaper agency. What we are trying to do in this paper is to use the dataset provvided and run a simple linear regression model to see the relevance effects between the advertising revenue of a newspaper agency and the quality of content that it produces.

## Data

The dataset we are going to use to reproduce the model is the historical records of the French daily newspaper and French television in the period from 1960 to 1974. The dataset contains annual data about the local and national newspapers agency, as well as data for television content. In 1967, the French government relaxed

its longstanding regulations against advertisement on television cable. This could be viewed as a substitute to advertisers who are using the daily newspaper to gain attentions.

The raw data contains 1196 observations with 52 variables. Each row contains data about a single year of values for a single newspaper agency. Mostly each newspaper agency have 15 rows, representing the time period of 1960 to 1974.

We can devide the dataset into 2 smaller dataset. A dataset for local newspaper and a dataset for national newspaper.

Some baseline characteristic of the dataset:

5 point summary on revenue on advertising for the local daily newspapers

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.       Max.      NA's
##    549717  12547410  30284258  66848228  98166316 416419200       125
```

5 point summary on revenue on advertising for the local daily newspapers

```
##      Min.   1st Qu.     Median       Mean   3rd Qu.       Max.      NA's
##   6683566  38249316  102705000  228134652 317868800 864369088        19
```

```
## # A tibble: 15 x 2
##     year 'total number of journalist working in national agency'
##    <int>                                                    <dbl>
##  1  1960                                                     1056
##  2  1961                                                     1102
##  3  1962                                                     1145
##  4  1963                                                     1199
##  5  1964                                                     1200
##  6  1965                                                     1215
##  7  1966                                                     1284
##  8  1967                                                     1308
##  9  1968                                                     1300
## 10  1969                                                     1298
## 11  1970                                                     1366
## 12  1971                                                     1396
## 13  1972                                                     1372
## 14  1973                                                     1230
## 15  1974                                                      963
```

```
## # A tibble: 15 x 2
##     year 'total number of journalist working in national agency'
##    <int>                                                    <dbl>
##  1  1960                                                     2236
##  2  1961                                                     2388
##  3  1962                                                     2515
##  4  1963                                                     2670
##  5  1964                                                     2892
##  6  1965                                                    3030.
##  7  1966                                                     3253
##  8  1967                                                     3354
##  9  1968                                                    3421.
## 10  1969                                                    3487.
## 11  1970                                                     3622
```

```
## 12   1971                                                      3787
## 13   1972                                                      3847
## 14   1973                                                      3894
## 15   1974                                                      4039
```

# Model

We know that the quality of content of a newspaper is positively correlated with the numbers of journalists working for that newspaper, because a newsroom with more people typically could get more stories written and proof. In the paper, the researchers point out that there could be a relationship between the quality of content on a newspaper and the total revenue of advertising. So, the model we chose to use is a simple linear regression model, with the response variable is the number of journalists working in each newspaper agency and the predictor variable is the total revenue from advertising. The total numbers of journalist in the model can be understood as how much work and content is put into the newspaper. Due to the dataset contains values for both local and national newspaper agency, we will separate the two.

After running linear regression on both type of newspaper, this is the model we got from the data:
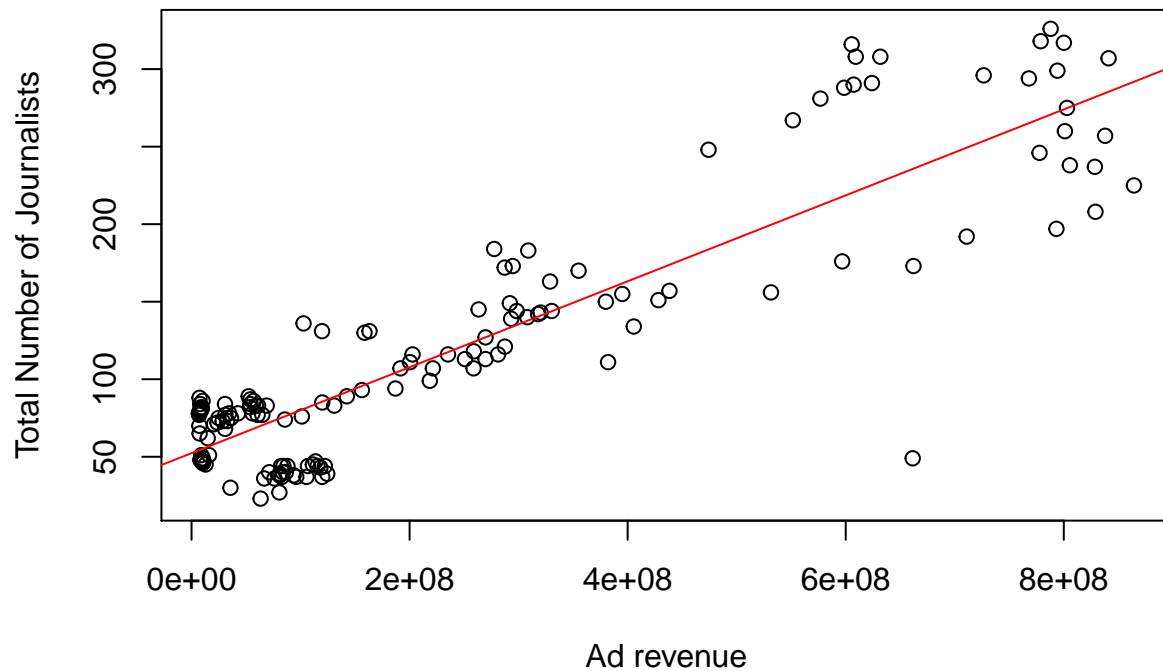
SLR for national newspaper:

$$number of journalist = 52.32 + 0,0000002771 * adrevenue$$

SLR for local newspaper:

$$number of journalist = 5.658 + 0.0000007012 * adrevenue$$

# Results

**Figure1. Scatterplot between ad revenue
and number of journalist working in a national newspaper agency**



```
## 
## Call:
## lm(formula = nb_journ ~ ra_cst, data = data_national)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -186.616  -29.915    0.876   22.023   95.889 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 5.232e+01  4.260e+00   12.28   <2e-16 ***
## ra_cst      2.771e-07  1.181e-08   23.46   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 37.49 on 144 degrees of freedom
##   (34 observations deleted due to missingness)
## Multiple R-squared:  0.7927, Adjusted R-squared:  0.7913 
## F-statistic: 550.6 on 1 and 144 DF,  p-value: < 2.2e-16
```

**Figure2. Scatterplot between ad revenue and
number of journalist working in a national newspaper agency**
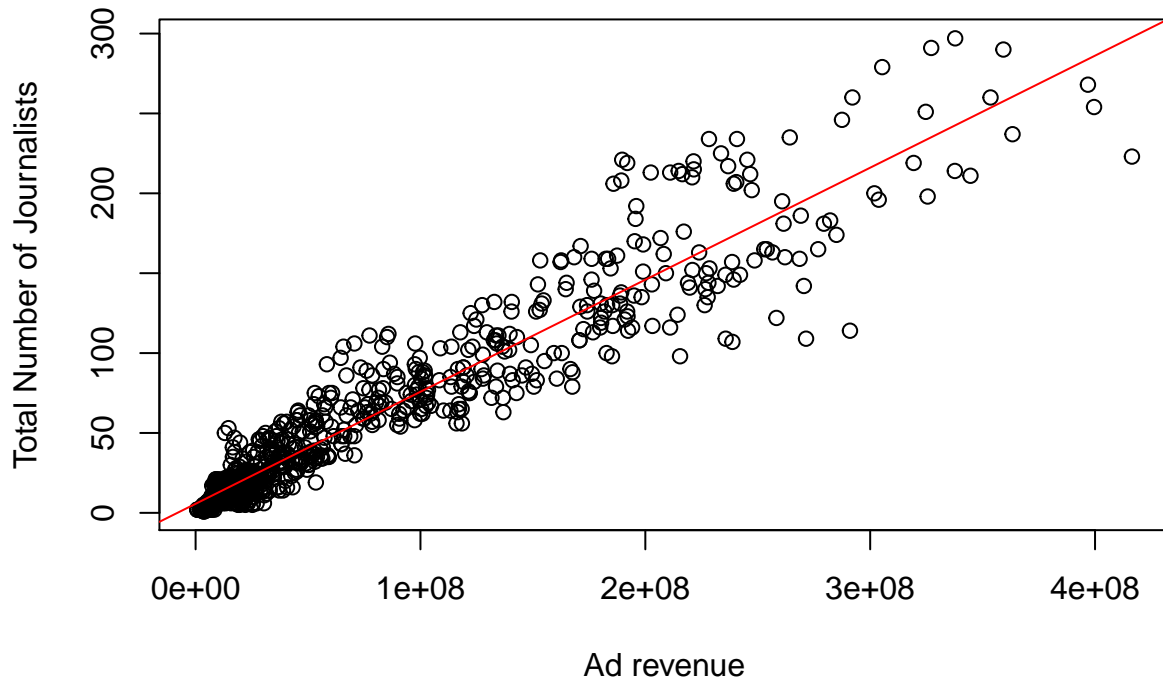


```
##
## Call:
## lm(formula = nb_journ ~ ra_cst, data = data_local)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -95.731  -7.297  -2.903   5.460  82.304
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 5.658e+00  8.416e-01   6.723  3.3e-11 ***
## ra_cst      7.012e-07  7.883e-09  88.954  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.34 on 835 degrees of freedom
##   (179 observations deleted due to missingness)
## Multiple R-squared:  0.9045, Adjusted R-squared:  0.9044
## F-statistic:  7913 on 1 and 835 DF,  p-value: < 2.2e-16
```

## Discussion

Looking at both of the graphs, we can already see that there is a positive linear relationship between the 2
variables.

By looking at the summary for both models, we see that the p-values for the intercept as well as the slope for both models are lower that the standard benchmark of 0.01. This mean that there are very strong evidence that we should reject the null hypothesis that there are no relationship between the revenue of a newspaper agency coming from advertising and its number of journalists. This mean there is a positive relationship between advertising revenue and the quality of the content published by the newspaper. We can also see that the R-squared for the model for national newspaper is 0.7927 and for local newspaper is 0.9045. R-squared represent of how much of the proportion of the dependent variable (number of journalist) can be explained by the explanatory variable (the ad revenue). Relatively these number are very high, indicating that the model fit quite well.

The purpose of the paper is to see the effects of a reduced revenue from advertising for a newspaper agency. One of the effect that we have explained through our model is advertising revenue is positively correlated with the number of journalist working in a agency, which coincide with the quality of content produced.

## Conclusion

With the relaxed law on advertising on television and the rise of usage in Internet in recent years, advertisers have more options to choose from. Consequently, an advertiser willingness to pay for an ad in newspaper has reduced, which resulted in a negative shock in the total revenue coming from advertisement for newspaper agency. This shock has indirectly reduced the quality of the content produced by the newspaper because of with the reduction of revenue, the size of the newsroom is also reduced.

## Weaknesses and next steps

One of the weakness of our model comes from the fact that there are observations that has the value of "NA", which has to be omit from the model. More data is benerfical as it can help us to produce a better model with better prediction.

Another weakness is that our model is a simple linear regression model. The next step could be to produce a model with multiple explanatory variables, which can help us understand better relationships between different variables.

## References

Angelucci, Charles, and Cagé, Julia. 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