Newspapers in Times of Low Advertising Revenues[*]

Reproduce the Paper: A Difference-in-Differences Analysis

Xi Cheng

December 9th, 2020

Abstract

In 2019, researchers Charles Angelucci and Julia Cagé have done the analysis of the relationship between the newspapers' content and the reduction in advertising revenues. They found robust evidence which demonstrated that a reduction in advertising revenues lowers newspapers' incentives to produce journalistic-intensive content through the difference-in-differences analysis. In this work, the same dataset was used and the difference-in-differences analysis was applied as well to reproduce their work. In general, similar results were obtained by this work compared to the published paper.

Keywords: Newspapers; Difference-in-Differences Analysis; Causal-Inference; Newspapers' Content; Advertising Revenues; Number of Journalists;

Introduction

broader context more detail about what you're interested in, what you did, what you found, why it's important

It has been argued that the information revolution is destroying the traditional newspaper industry around the world right now, from local newspaper to national newspaper. It has been demonstrated that, with the rise of the Internet, the revenues, the advertisers and even the number of journalists employed of the newspaper industry has been steadily decreased, roughly since 2005 (). Therefore, there is a growing concern that the quantity of high-quality content might decrease as well (). However, the direct accurate causes of the declination of the newspaper remain debatable. Although it is obvious that there is a strong negative correlation between the trend of the Internet and the trend of traditional newspaper industry. One could still argue that, there might exist a third factor such as the changing customer preferences driving both trends (). Similarly, the causality between the decrease in number of journalists employed and the decrease in the advertising revenues remains uncertain as well. Hence, specific hypothesis must be proposed.

the research

In their study, they focused on the effects of the decline in advertising revenues on the quality of the newspapers' content (). To investigate this relationship, they built a model based on several conditions and assumptions, including using a monopoly newspaper which is able to choose the prices of subscription and advertisement, letting readers to be heterogeneous in the types of high-quality content, assuming that more journalist-intensive content increases more costs. In terms of the statistical method, the difference-in-differences analysis was applied to derive the causality (). With all above framework, they was trying to prove that a decline in advertising revenues may cause a decline in the amount of high-quality content

[*] Code are available at: https://github.com/

produced, a drop in reader subscription prices and a change in the composition of readership toward a less wealthy readership ().

the outline and summary results

Methods

Data

The characteristics of data through tables (Table 1. and Table 2.).

Table. 1: Characteristics Summary of National Daily Newspapers

	Overall (N=181)
Unit buyer price	
N-Miss	29
Mean (SD)	3.592 (1.255)
Range	2.395 - 9.345
Subscription price	
N-Miss	33
Mean (SD)	2.807 (0.739)
Range	1.925 - 5.630
Ad rate (listed price)	
N-Miss	60
Mean (SD)	121.135 (80.959)
Range	17.535 - 274.200
Total revenues (€)	
N-Miss	19
Mean (SD)	424970741.802 (403009767.658)
Range	18918480.000 - 1482414336.000
Revenues from advertising (€)	
N-Miss	20
Mean (SD)	228134652.270 (257993642.079)
Range	6683565.500 - 864369088.000
Number of journalists	
N-Miss	23
Mean (SD)	116.671 (80.562)
Range	21.000 - 326.000
Newshole (nonadvertising space)	
N-Miss	43
Mean (SD)	13.198 (4.082)
Range	6.320 - 24.657

Table. 2: Characteristics Summary of Local Daily Newspapers

	Overall (N=1016)
Unit buyer price	
N-Miss	105
Mean (SD)	3.175 (0.790)
Range	0.818 - 5.700
Subscription price	

	Overall (N=1016)
N-Miss	120
Mean (SD)	2.770 (0.735)
Range	0.682 - 4.687
Ad rate (listed price)	
N-Miss	328
Mean (SD)	80.333 (72.623)
Range	3.757 - 327.200
Total revenues (€)	
N-Miss	128
Mean (SD)	145562770.985 (176075111.269)
Range	809563.625 - 1025858560.000
Revenues from advertising (€)	
N-Miss	125
Mean (SD)	66848228.204 (79245513.189)
Range	549717.250 - 416419200.000
Number of journalists	
N-Miss	109
Mean (SD)	53.401 (57.895)
Range	1.000 - 297.000
Newshole (nonadvertising space)	
N-Miss	108
Mean (SD)	12.345 (3.932)
Range	1.860 - 34.435

Model

Model Details

The complete model was shown here:

$$Revenue \sim Normal(\frac{1}{1 + exp(-(a + b_i x_i +))}) \tag{1}$$

where the a is the intercept, b representing coefficients of different variables. Particularly, b_i and x_i

Equation (1) represents the complete model, and Equation (2) represents our final model, which did not include employment status.

Model is approriate

Discussion on features selection.

All work were done in `R` (version 4.0.2) (R Core Team 2020) and `Rstudio` (version 1.3.1093). `Tidyverse` (version 1.3.0) was used for data wrangling and visualization (Wickham et al. 2019). R package `forcats` (version 0.5.0) was also used for data pre-processing (Wickham 2020). There are other packages used such as `captioner`, `gridExtra`, `broom`, `Haven`, `magrittr`, `knitr`, `labelled` and `arsenal` (Alathea 2015; Hlavac 2018; Heinzen et al. 2020; Xie 2020; Wickham and Miller 2020; Auguie 2017; Robinson, Hayes, and Couch 2020; Bache and Wickham 2014; Larmarange 2020). Code are available at: https://github.com/.

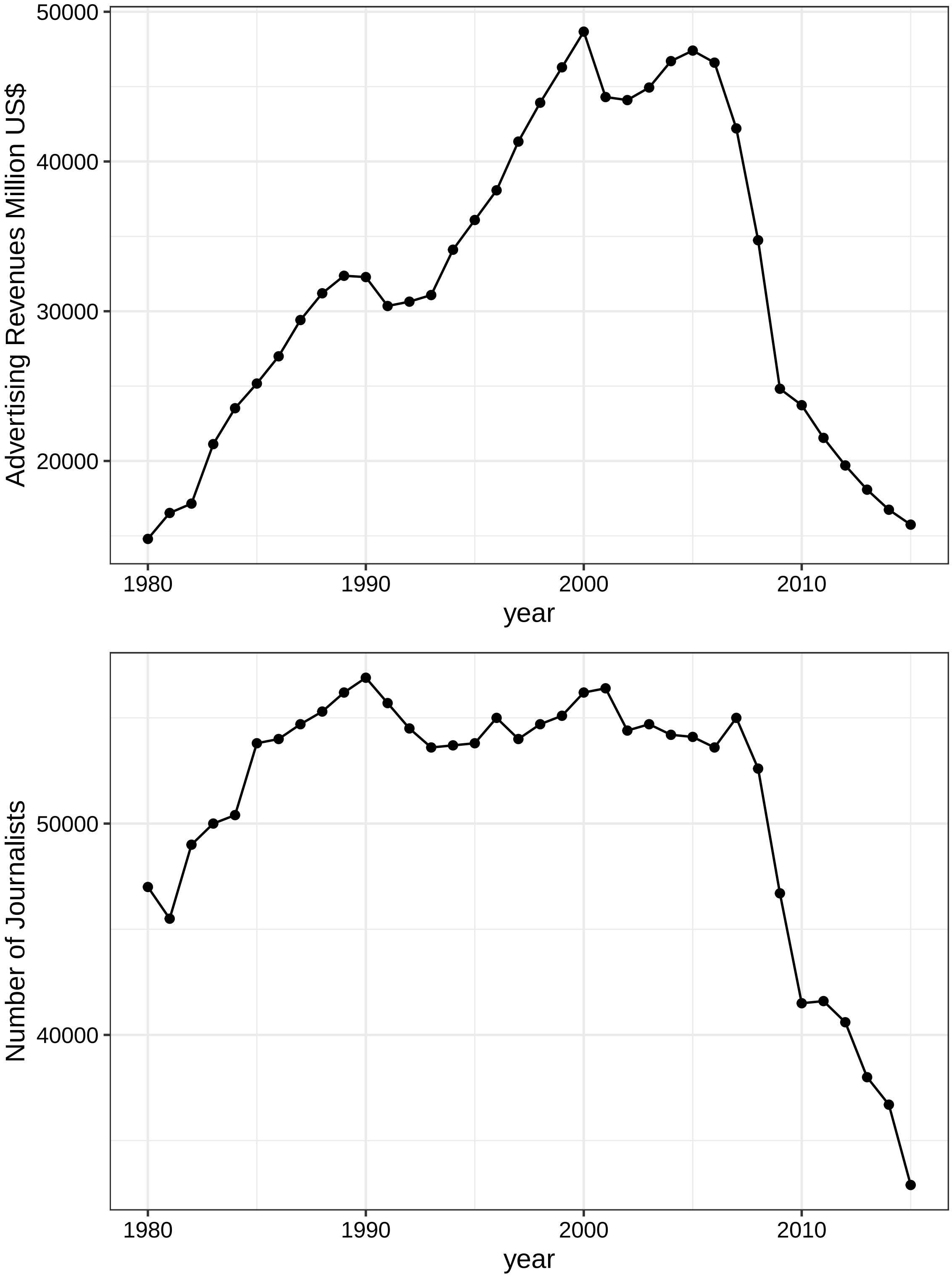

Figure 1: Newspaper Advertising Revenues (in dollars) and Number of Journalists in the United States, 1980-2015

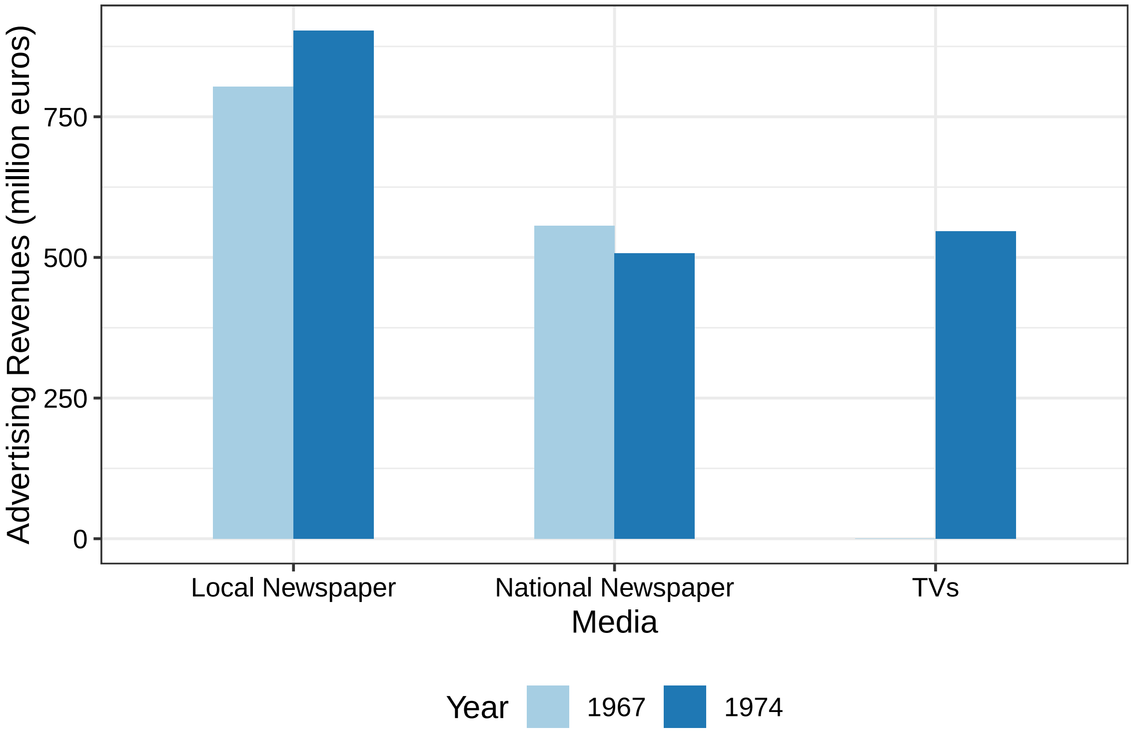

Figure 2: Advertising Revenues in France by Media Outlets

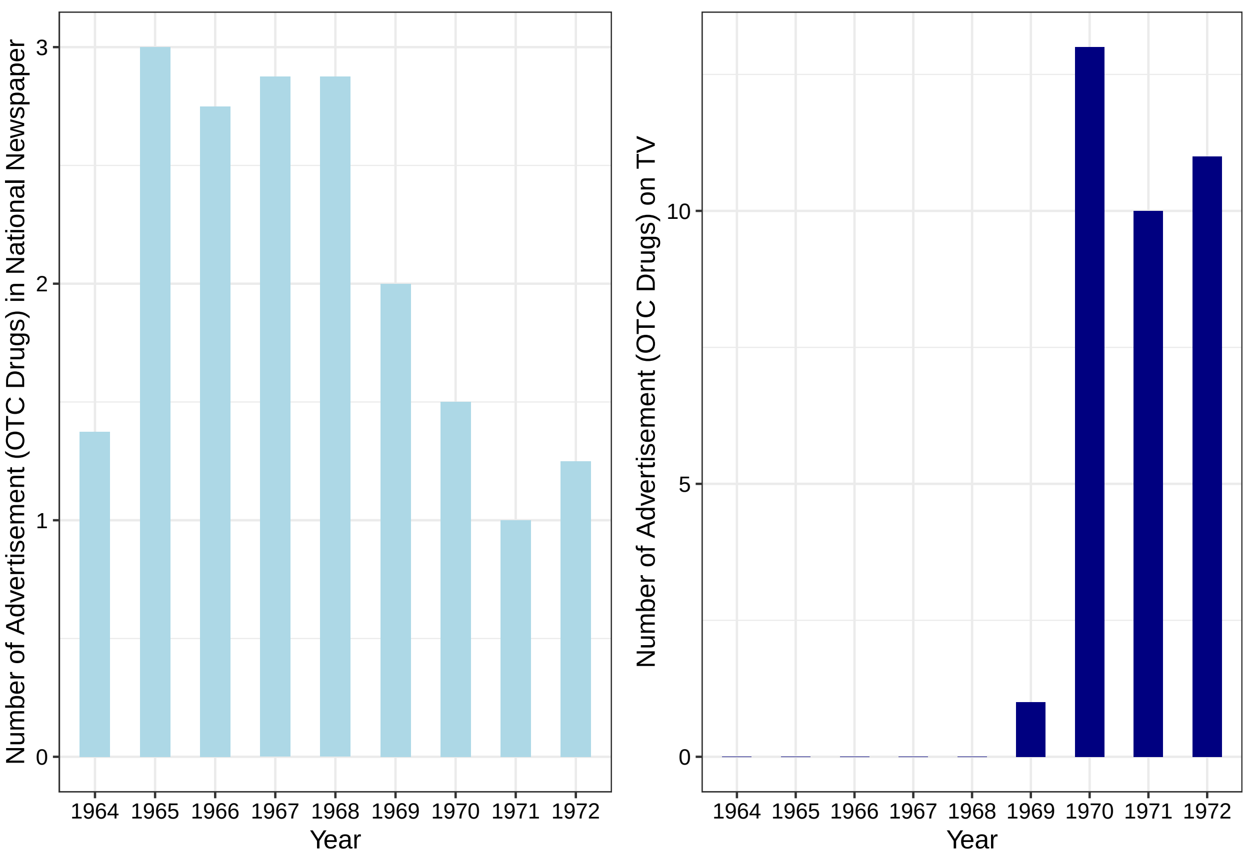

Figure 3: Effects on National Newspaper from Advertisment on TV

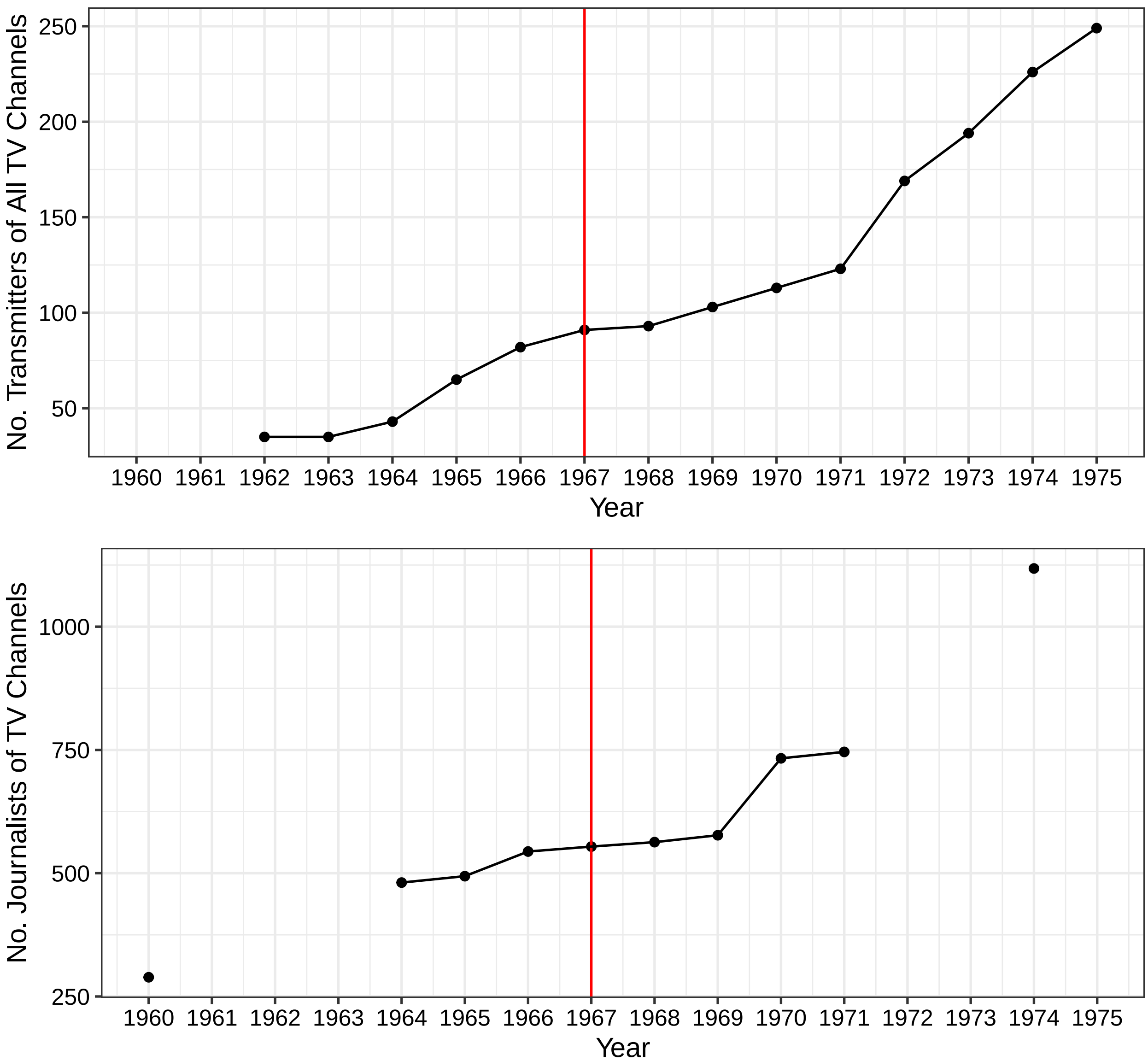

Figure 4: Measure of Television Quality

Results

Table. 3: Advertising Side

Table 3:

	Dependent variable:			
	Ad Revenues	Ad Revenues (per)	Ad Listed Price	Ad Space
	(1)	(2)	(3)	(4)
National × Post-TV ad	−0.239	−0.146	−0.395	−0.030
	t = −7.596***	t = −5.284***	t = −5.408***	t = −0.540
Intercept	17.867	6.251	3.874	1.874
	t = 339.540***	t = 134.804***	t = 28.439***	t = 21.516***
Observations	1,052	1,051	809	1,046
R^2	0.986	0.911	0.897	0.869
Adjusted R^2	0.984	0.902	0.885	0.857
Note:				*p<0.1; **p<0.05; ***p<0.01

Table. 4: Reader Side

Table 4:

	Dependent variable:				
	Sub. Price	Unit Price	Circulation	Share of Sub.	Revenues (sales)
	(1)	(2)	(3)	(4)	(5)
National × Post-TV ad	−0.111	0.004	−0.075	0.234	−0.133
	t = −8.337***	t = 0.346	t = −3.515***	t = 7.214***	t = −5.252***
Intercept	0.809	0.954	11.644	4.633	18.145
	t = 37.147***	t = 49.746***	t = 326.374***	t = 87.372***	t = 420.562***
Observations	1,044	1,044	1,070	1,044	1,046
R^2	0.943	0.953	0.991	0.974	0.991
Adjusted R^2	0.938	0.948	0.991	0.971	0.990
Note:					*p<0.1; **p<0.05; ***p<0.01

Table. 5: Quality

Discussion

Interpretation of results

What have we learnt from the model

Detailed information of the final model fitting can be found in the **Appendix**.

Table 5:

	Dependent variable:			
	No. Journalists	Average Payroll	No. Pages	Newshole
	(1)	(2)	(3)	(4)
National × Post-TV ad	−0.211	0.057	−0.026	−0.045
	t = −6.958***	t = 1.137	t = −1.229	t = −1.986**
Intercept	3.828	7.571	3.039	2.631
	t = 77.314***	t = 104.484***	t = 93.916***	t = 75.182***
Observations	1,046	723	1,046	1,046
R^2	0.979	0.567	0.933	0.902
Adjusted R^2	0.977	0.524	0.926	0.893
Note:				*p<0.1; **p<0.05; ***p<0.01

All statistical modeling has two frames: the small world of the model itself and the large world we hope to deploy the model in.

The sex and gender problem in modern survey

It is worthwhile mentioning that

Weaknesses and next steps

As we mentioned above,

Appendix

References

Alathea, Letaw. 2015. *Captioner: Numbers Figures and Creates Simple Captions*. https://CRAN.R-project.org/package=captioner.

Auguie, Baptiste. 2017. *GridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R*. https://CRAN.R-project.org/package=magrittr.

Heinzen, Ethan, Jason Sinnwell, Elizabeth Atkinson, Tina Gunderson, and Gregory Dougherty. 2020. *Arsenal: An Arsenal of 'R' Functions for Large-Scale Statisticalsummaries*. https://CRAN.R-project.org/package=arsenal.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Larmarange, Joseph. 2020. *Labelled: Manipulating Labelled Data*. https://CRAN.R-project.org/package=labelled.

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

Robinson, David, Alex Hayes, and Simon Couch. 2020. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Wickham, Hadley. 2020. *Forcats: Tools for Working with Categorical Variables (Factors)*. https://CRAN.R-project.org/package=forcats.

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.

Xie, Yihui. 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*.