What is Affecting Total Loss Percentage of Television Audience: An empirical study on series Blindspot *

Wanlin Ji†

June 2017

Abstract

It is crucial for media to understand how many people are watching a show and its dynamic pattern, directly influencing their investments with advertisements. Total loss percentage indicates the losing pattern of audience. This empirical study is based on the minute-by-minute rating data from acclaimed television series Blindspot, ranging from September 2015 to May 2016. The results showed that commercial has a significant influence on the total loss percentage, and the time affects the total loss percentage through rating.

keywords: Rating, advertising, audience loss.

JEL classification: C36, D12, M37

*I thank Richard W. Evans, Benjamin Soltoff and Ging Cee Ng for their instructions and suggestions, as well as participants on computational poster presentation for their helpful comments.

†University of Chicago, Master of Arts Program in Computational Social Sciences, (562) 298-0258, wji@uchicago.edu.

Contents

1 Introduction **2**

2 Data **2**
2.1 Nielsen Minute-by-minute Television Rating Dataset 3
2.2 Variables 3
2.3 Exploratory analysis 4

3 Model **5**
3.1 Hypothesis 6
3.2 Instrumental Variable 6
3.3 2SLS model 7

4 Result **8**

5 Conclusion **8**

A-1 Exploratory figures **10**

A-2 Variable distribution **10**

1 Introduction

It is crucial for media to understand how many people are watching a show and its dynamic pattern, directly influencing their investments with advertisements.

Past literatures on minute-by-minute TV rating dataset has been focused on the influencing mechanism of rating. Alavy et. al (2010) Van Meurs (1998) adapted ratings data to predict the demand of football programs, while Harrington (2012) integrated the panel measurement with online twitter data to analyze when people are distracted from TV. However, few has discovered the dynamics of audience loss by minute.

This study is based on the case of TV series Blindspot, aired from 9/21/15 - 5/23/16. The Nielsen Rating dataset features the minute-by-minute property of major variables, which allows us to look deeper into the dynamics of audience behavior when the TV series are on air and find out how these factors are influencing audience loss.

2 Data

Our research is based on the case of acclaimed series Blindspot. Blindspot is an American crime drama television series created by Martin Gero, starring Sullivan Stapleton and Jaimie Alexander. The series is aired on DKNY network from September 21st, 2015 to May 23rd, 2016. A back nine order was given on October 9, 2015, bringing the first season to a total of 22 episodes, plus an additional episode bringing the order to 23 episodes. Blindspot focused on a beautiful woman named Jane Doe left in Times Square inside of a duffel bag with no memory and her body completely covered in tattoos. The main character Kurt Weller, and his FBI Team are to investigate the circumstances around Jane and solve the puzzles hidden in her intricate tattoos.

2.1 Nielsen Minute-by-minute Television Rating Dataset

Our primary data source for Blindspot is Nielsen Minute-by-minute Television Rating Dataset. Nielsen Rating Data provides a detailed and minute-by-minute measurement of ratings and total loss percentage of audience for every one of 23 episodes of Blindspot" through all available television networks, DKN in this case. Nielsen sets up previously electronic and proprietary metering equipments to carry out their audience measurement throughout the country. Nielsen also cooperates with television network providers to collect audience behavior data and capture information about whats being viewed and when. The minute-by-minute precision of data extends our ability for analyzing each episode dynamically to see if there were significant points in a given episode that resulted in losing large chunks of the audience.

The Nielsen Nielsen Minute-by-minute Network Ratings Dataset has some limitations for analysis. Generally speaking, the quality of the content has a strong influence on the audience behavior, and this information is highly abstract for the rating and audience loss. Panel measurement does not quantify the narratives. This could cause the endogeneity problem between ratings and error term, because of the omitted variable about narrative quality. We would discuss how to overcome endogeneity further in our model section.

2.2 Variables

There are nine variables in the "Blindspot" minute-by-minute ratings data. According to the Time variable, all the factors are measured by minute. Variables include:

- X: Index order for data point in each minute among all 23 episodes
- Network: The name of the network the show (also known as a telecast) aired on
- Date: The airing date of a particular telecast
- Time: The local time indicating the particular minute for the show
- Program: The name of the telecast
- Length: The duration length of the telecast
- Rating: the rating of the show, i.e. share of people/households watching the

show relative to the total people/households with a TV

- Minute in Commercial: Dummy variable for whether there is a commercial during that given minute or not (1 = minutes with commercials)

- Total Loss Percentage: Percentage of individuals who stopped watching the broadcast in that minute from the total number of viewers in that minute

Through the inspection, it can be shown that the Length, Network, Program are almost the same for all the cases in dataset. Our focus would be on the main four variables available to us: Total Loss Percentage, X, Rating, Minute In Commercial, with the Total Loss Percentage as the dependent variable.

2.3 Exploratory analysis

After inspection, there is no missing value in the dataset, thus no need for data imputation. Next we want to carry out exploratory analysis on our data at hand. This is helpful to better understand the data and to test the hypotheses based on data experience. However, sometimes there are counter-intuitive findings so it is important to have an unbiased understanding of the data before proceeding with more advanced analysis. Firstly, the five number summary for key variables can be found in the table below.

Table 1: Descriptive statistics for key variables

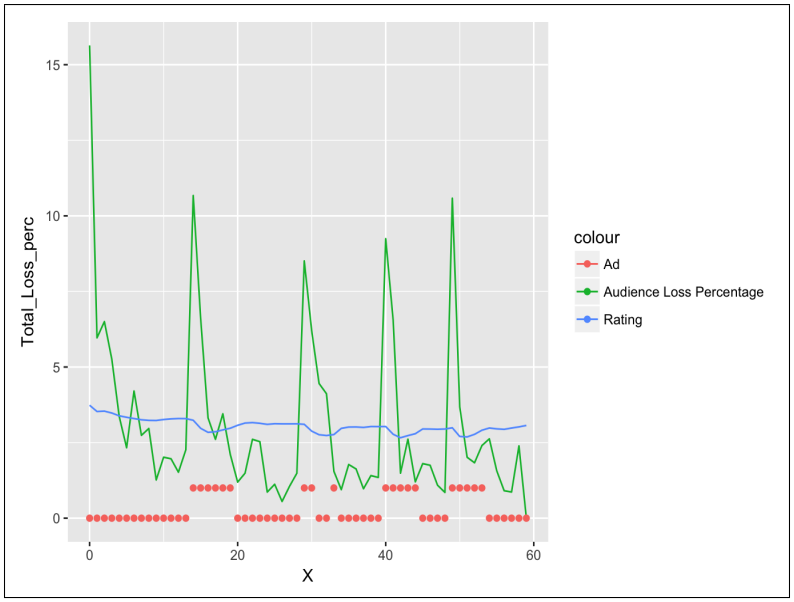
Variables	Min	1st Quartile	Median	3rd Quartile	Max
X	0	521.8	1043.5	1565.2	2087.0
Rating	0.955	1.372	1.552	1.964	3.831
Minute In Commercial	0	0	0	1	1
Loss Percentage	0	1.776	2.548	3.824	29.634

* Calculated under R. All codes disclosed on the Github.

To identify the influence of rating and commercial factors on total loss percentage, we are concerned with the pattern of loss in long term. Appendix A-1 figure shows the overall tendency of total loss percentage during the year when Blindspot was on air. Although we can tell the loss percentage shows fluctuations around the same level, it is apparent that data points are highly condensed in the long run. Then we

zoom into the first episode of Blindspot to observe its short-run relationships between major variables.

Figure 1: Variable tendency in the first episode



By comparing the three major variables, commercial, rating and total loss percentage, we found that the scope can really mislead us. The ratings are not so violent as we saw previously. Until now, we have got some solid knowledge of our data at hand.

Through above graphs, we find that the total loss percentage has high correlation with the ratings. The ratings show similar patterns of fluctuations in the short term, reflecting the endogeneity problem. Since the total loss percentage and rating are influencing each other in our dataset and the total loss percentage is our dependent variable, an instrumental variable that is correlated with rating but not with the total loss percentage can clarify the causal relationship between these two.

3 Model

Because total loss percentage and rating have endogeneity problem, we use X as the instrument variable to rating, where X is the time order of the data by minute

(partitioned episode).

3.1 Hypothesis

Now let's consider some of the hypothesis. We must pick up a dependent variable that could help us predict. From my perspective, the industry are more concerned with the number of audience, so we should think of the total loss percentage as our primary target here. Note the number of viewers who join the watching during the show is not available to the research, so it is impossible to estimate the real number of audience that are leaving this show, but we can still aim to manage the viewers loss according to the insight from marketing that it is always more expensive to earn a new customer than to prevent losing an old customer. Same goes with viewers.

Next we want to consider the independent variables that could have an effect on the total loss percentage to find the primary drive behind it. We want to assume the rating stands for the quality of narratives, then we can see the rating an explanatory variable for analyzing total loss percentage. This seems plausible explanation. But we also may pay attention to the sudden fluctuations that comes with the total loss percentage, which is largely corresponding to Minute In Commercial.

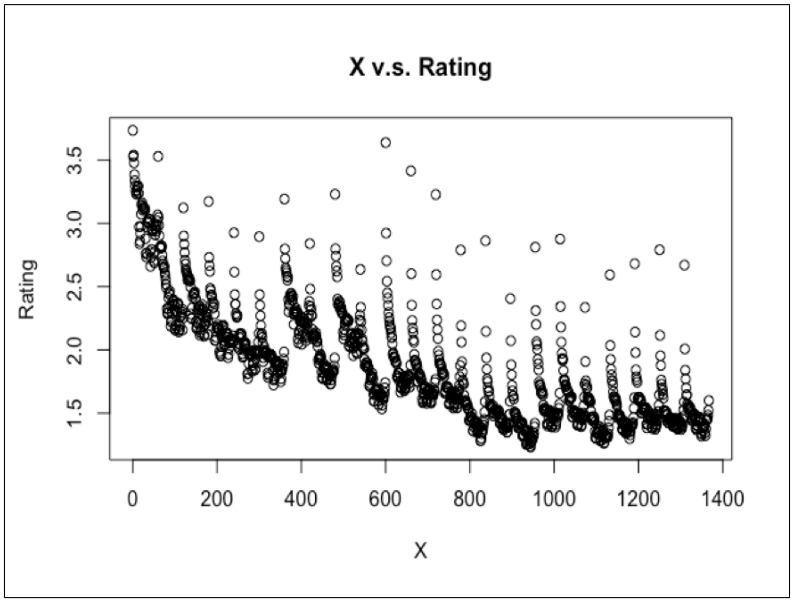
3.2 Instrumental Variable

The endogeneity problem lies in the correlation between our response variable total loss percentage and rating. An instrumental variable that is related to rating but not to total loss percentage can help reduce the correlation between rating and error terms.

The time order is chosen as the instrumental variable. Firstly, the time order has a significant relationship with rating. This tendency can be shown in the figure below, we can find the rating is decreasing as time goes further.

And we found that the time order does not have a relationship with total loss percentage, indicating the percentage of people leaving this show is not affected by

Figure 2: Relationship between rating and time order



the time order.

The relationships between time and rating/loss are given below, showing different tendencies in the long run.

3.3 2SLS model

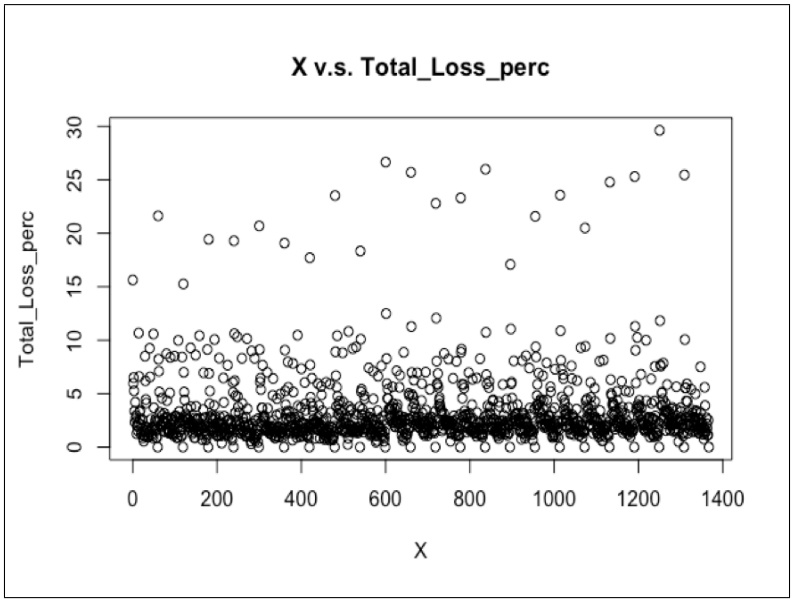
First stage:

$$Rating = \alpha + \beta_1 X + \beta_2 MinuteInCommercial + \beta_3 X * MinuteInCommercial \quad (1)$$

Second stage:

$$TotalLossPercentage = \theta + \pi_1 Rating + \pi_2 MinuteInCommercial + \pi_3 X * MinuteInCommercial \quad (2)$$

Figure 3: Relationship between total loss percentage and time order



4 Result

The results of first and second stage regression is shown below:

Table 2: First stage estimation

Coefficients	Estimate	Std. Error	t-value
Intercept	2.584***	0.02	1043.5
1565.2	2087.0		
X	0.955	-9.6e-04***	1.552
1.964	3.831		
Commercial	-0.22	0	0
1	1		
Interaction Term	1.338e-04**	1.776	2.548
3.824	29.634		

* Calculated under R. All codes disclosed on the Github.

5 Conclusion

Put conclusion here.

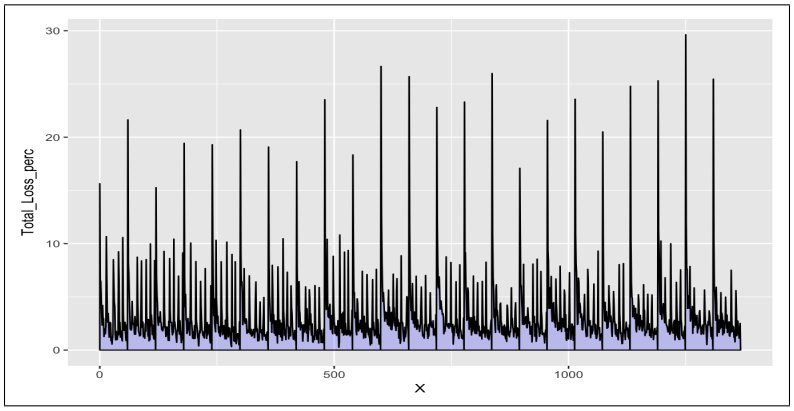
References

Meurs, L. Van, "Zapp! A study on switching behavior during commercial breaks," *Journal of Advertising Research*, 1998, *38* (1), 43–44.

APPENDIX

A-1 Exploratory figures

Figure 4: Dependent Variable: Audience loss percent tendency



A-2 Variable distribution

Figure 5: Distribution of Rating

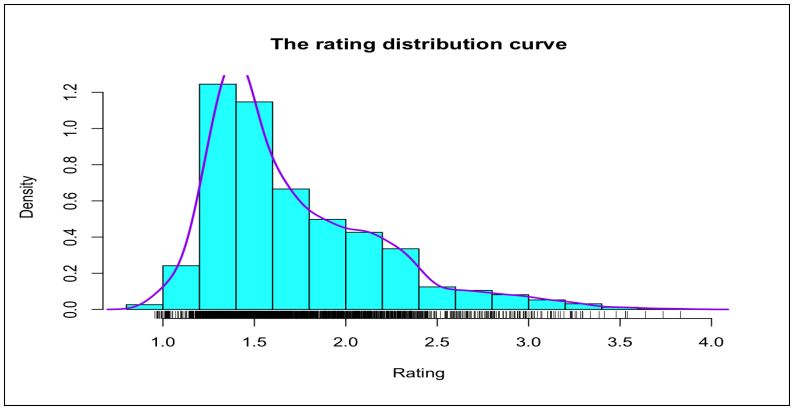


Figure 6: Distribution of Total Loss Percentage