Department of Economics- FEA/USP

Is *Rotten Tomatoes* killing the movie industry? A regression discontinuity approach

Marislei Nishijima
Mauro Rodrigues
Thaís L. D. Souza

Working Paper Series Nº 2021-12

DEPARTMENT OF ECONOMICS, FEA-USP
WORKING PAPER Nº 2021-12

Is *Rotten Tomatoes* killing the movie industry? A regression discontinuity approach

Marislei Nishijima (marislei@usp.br)

Mauro Rodrigues (mrodrigues@usp.br)

Thaís Luiza Donega Souza (Thais.donega@usp.br)

Abstract:

We study the effect of expert reviews on consumer choice, focusing on the movie industry. Specifically, we estimate the impact of a popular movie review aggregator – Rotten Tomatoes (RT) – on box office revenue, using a sample of 1,239 movies widely-released in the U.S. between 1999 and 2019. RT's rating system allows us to use regression discontinuity, thus avoiding problems associated with omitted unobservable characteristics of movies (such as quality and commercial appeal). We do not find evidence that RT ratings affect box office performance.

Keywords: Expert reviews; Rotten Tomatoes; Movie industry; Films; Regression discontinuity.

JEL Codes: D8; D9; C21; C39.

Is *Rotten Tomatoes* killing the movie industry? A regression discontinuity approach

Marislei Nishijima
Institute of International Relations - University of Sao Paulo
Sao Paulo, Brazil
marislei@usp.br
+55 11 3091-0526
ORCID: 0000-0003-1162-7987

Mauro Rodrigues
Department of Economics, University of Sao Paulo
Sao Paulo, Brazil
mrodrigues@usp.br
ORCID: 0000-0003-3745-3989

Thaís Luiza Donega Souza
School of Arts, Sciences and Humanities - University of Sao Paulo
Sao Paulo, Brazil
Thais.donega@usp.br
ORCID: 0000-0002-2879-694X

Abstract: We study the effect of expert reviews on consumer choice, focusing on the movie industry. Specifically, we estimate the impact of a popular movie review aggregator – Rotten Tomatoes (RT) – on box office revenue, using a sample of 1,239 movies widely-released in the U.S. between 1999 and 2019. RT's rating system allows us to use regression discontinuity, thus avoiding problems associated with omitted unobservable characteristics of movies (such as quality and commercial appeal). We do not find evidence that RT ratings affect box office performance.

Key-words: expert reviews, Rotten Tomatoes, movie industry, films, regression discontinuity

JEL Codes: D8, D9, C21, C39

We gratefully acknowledge the financial support for this project from the FAPESP, CAPES, and CNPq agencies.

Is *Rotten Tomatoes* killing the movie industry? A regression discontinuity approach

1. Introduction

Consumers may resort to experts' opinions when the choice to buy a product involves uncertainty about quality. This is the case of experience goods, such as movies (Reinstein and Snyder 2005). With the expansion of the internet, the availability of movie reviews proliferated. Consumers may have difficulty processing all this information, and, as a result, aggregators emerged. We study the impact of a popular aggregator – Rotten Tomatoes (RT) – on movie consumption.

RT gathers information on reviews from several professional critics and rates them as positive or negative. It then reports a score, which is the percentage of positive reviews a movie received (the "Tomatometer"). Additionally, it classifies movies into two categories: "fresh" if a film has at least 60% positive reviews, and "rotten" otherwise. RT's importance is recognized throughout the industry. Movie executives often blame RT for the demise of certain productions because of bad scores assigned by the aggregator (see, for instance, New York Times 2017). [1]

Nonetheless, a positive correlation between RT scores and box office revenues does not imply causation because of unobservable quality or commercial appeal. Bad movies tend to both get bad reviews and perform poorly at the box office. RT's dichotomic classification ("fresh" and "rotten") allows us to sidestep this problem through regression discontinuity techniques. We thus compare the box office performance of movies that are close to the 60% cutoff. They have similar RT scores, but only some of them received the "fresh" label.

Our sample comprises movies released in the U.S. market between 1999 and 2019. Industry complaints about the aggregator revolve around larger scale productions – for instance, a

[1] https://www.nytimes.com/2017/09/07/business/media/rotten-tomatoes-box-office.html?searchResultPosition=3

blockbuster that flops allegedly because of bad reviews. For this reason, we focus on wide releases. Besides, we have more information on these films than on limited releases. We also evaluate another (more stringent) discrete classification from RT, which labels movies with at least 75% of positive reviews as "Certified Fresh". We do not find evidence of a robust impact of RT on box office in either threshold.

Our work is part of a literature that studies the impact of expert opinions on consumer choice (see Ginsburgh, 2003, for a broad discussion). It is more closely related to papers that investigate movie critics' role (Eliashberg and Shugan, 1997; Reinstein and Snyder, 2005; Brown et al., 2012; Thrane, 2018). There are, however, fewer papers that focus on aggregators such as RT (an exception is Basuroy et al., 2020). We are unaware of other research that uses RT's discrete classification to understand the impact of expert opinions on movies' commercial success.

2. Data and Empirical Strategy

As mentioned, the "Tomatometer" indicates the percentage of positive reviews received by a movie. Moreover, RT categorizes movies as "fresh" or "rotten", following a rule based on a discontinuity at 60%. RT employs another threshold reserved for films with a substantial share of positive reviews (75% or higher), which are categorized as "certified fresh". However, to receive this label, a movie has to have a minimum number of reviews (for wide releases, at least 80 reviews in total, and at least 5 from top critics). Figure 1 provides examples from RT's website. The labels "rotten", "fresh" and "certified fresh" are quite salient, as one can see.

We expect critics' opinions to have a stronger influence in the first days after the movie's release. As time passes, other factors tend to play a role in moviegoers' choices, such as word of mouth (Eliashberg and Shugan, 1997). For this reason, we focus on box office performance during opening weekend. For each movie in our sample, we reconstruct RT's score right before release, using reviews available until opening day (individual reviews and their dates are available at RT's website). As previously discussed, we consider only wide releases, that is, movies screened in at least 600 theaters throughout the U.S. during opening weekend. This is due to our motivation (discussed in the Introduction) and to data availability issues – for most limited releases we could not calculate the RT score before opening, as they had an insufficient number of reviews or no reviews. Box office data are in constant U.S. dollars and come from boxofficemojo.com. This gives us a set of 1,239 movies released in the U.S. between 1999 and 2019.

Following Calonico et al. (2016) and Cattaneo et al. (2020), we estimate local linear regressions within a bandwidth centered around each threshold. For instance, around the 60% threshold, we have a set of movies that have similar RT scores, but only some of them are classified as "fresh". For these movies, we estimate equation (1):

$$Y_i = \alpha + \tau . T_i + \beta_1 \cdot (X_i - 0.6) + \gamma \cdot T_i \cdot (X_i - 0.6) + \boldsymbol{Z_i}'\boldsymbol{\theta} + \varepsilon_i \quad (1)$$

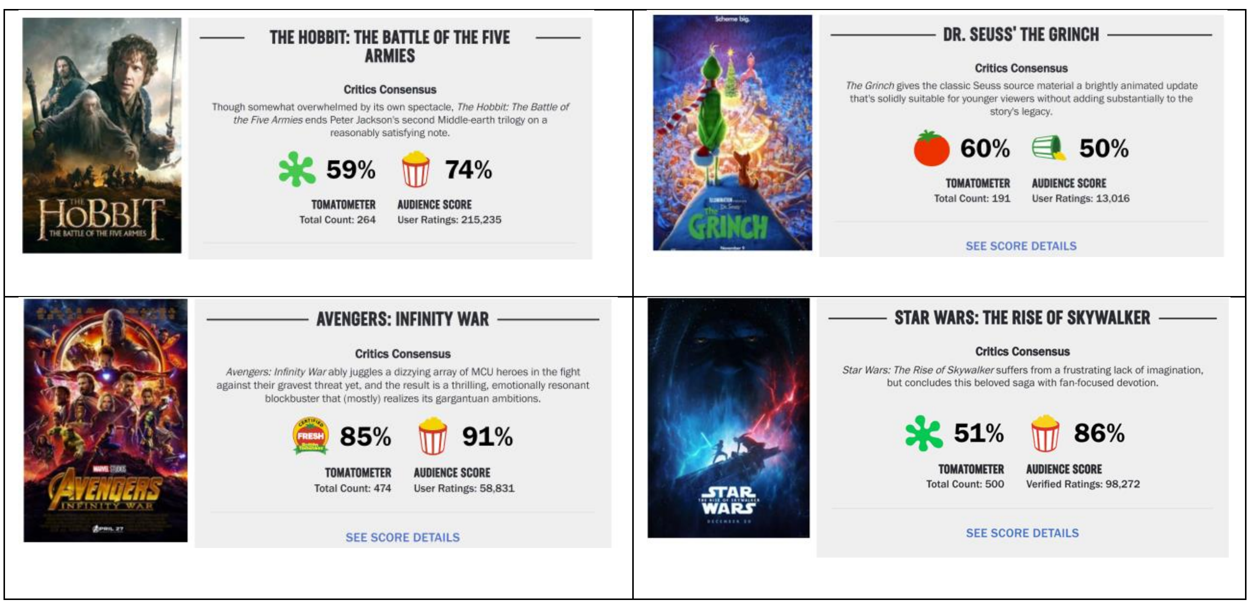
where Y_i is the logarithm of the opening revenue of film *i*, and X_i is *i*'s RT score (share of positive reviews) before its release. Moreover, T_i is equal to 1 if film *i* is classified as "fresh" (that is, if $X_i \geq 60\%$), and 0 if "rotten" (if $X_i < 60\%$); $\boldsymbol{Z_i}$ is a vector of control variables, which encompasses only information available before the film release date. Controls, obtained at Box Office Mojo website, include the log of the number of theaters a movie is shown at its release, log of film's budget, along with *dummies* for genre, big distributor, MPAA ratings, release year, and release month. Information on budget is limited to a subset of movies (about 77% of our sample). We thus run regressions with and without this variable.

We are interested in the parameter τ, which measures the effect of moving from the "rotten" to the "fresh" category on box office revenue. We undertake a similar procedure using the 75%

threshold, with τ capturing the impact of moving from "fresh" to "certified fresh". In this case, however, the RT score does not determine uniquely whether a movie would be classified as "certified fresh" since it also needs to have a minimum number of reviews. Therefore, for this cutoff, we use the fuzzy regression discontinuity approach, where the variable $1\{X_i \geq 75\%\}$ serves as an instrument for $1\{\text{If movie } i \text{ is "certified fresh"}\}$.

Following Cattaneo et al. (2020), the estimation procedure sets the bandwidth optimally, taking into account the tradeoff between bias and variance associated with the length of the interval considered. Observations within the bandwidth are weighted according to a triangular kernel centered around the cutoff. This implies that observations closer to the cutoff receive more weight in the estimation.

Figure 1 – Examples of movies featured in RT's website

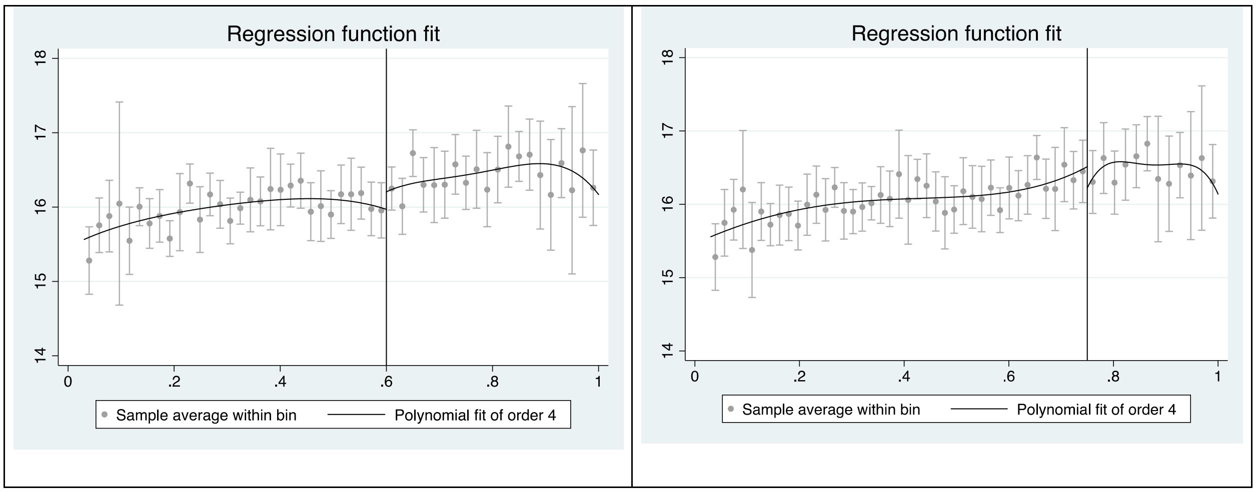


Source: Rotten Tomatoes website

3. Results

To get a sense of possible discontinuities in the data, we first show plots of the log of box office opening revenue against the RT score, using all the movies in our sample. We fit polynomials of fourth degree both to the left and to the right of each threshold to check for “jumps”. Polynomials of degree 2, 3, or 5 yield similar results. Figure 2 shows the result of this exercise. We can see a positive “jump” at the 60% threshold (left panel). In the 75% threshold (right panel) the estimated discontinuity is negative, but its magnitude is much smaller. As we shall see next, neither jump is significant when we estimate local linear regressions.

Figure 2 - Logarithm of opening revenue against RT score



These results are displayed in Table 1. The dependent variable is the log of box office revenue at the opening weekend. The upper panel shows estimation outputs of equation (1) for the 60% threshold, using data in a bandwidth around the cutoff. We focus on our parameter of interest (τ), which represents the jump at the threshold. Column (1) reports estimates for a regression with no controls. In column (2) we add the controls described in the last section, except for film’s budget. Column (3) shows estimates for the model with all controls, including budget. In all cases, the estimated coefficient is not statistically distinguishable from zero at the conventional levels of significance.

We finally evaluate if the results are different for a more recent sample - films released from 2010 to 2019. RT’s influence may have changed over time, possibly because of the expansion of the internet and social media. Column (4) shows that the results are insignificant for this subsample as well.

Table 1 – Estimates at 60% and 75% thresholds based on Rotten Tomatoes classification

VARIABLES	Without controls (1)	With controls (2)	With controls + budget (3)	With controls + budget and >2010 (4)
		60% Sharp RD		
τ	0.1708	0.0088	0.0616	0.1834
	(0.219)	(0.180)	(0.165)	(0.198)
Eff. Obs. Left/Right	250/225	196/174	129/130	43/46
Obs. Left/Right	770/469	770/469	573/378	206/165
		75% Fuzzy RD		
τ	0.0580	0.4753	0.2601	-0.0021
	(1.022)	(0.482)	(0.400)	(0.215)
Eff. Obs. Left/Right	99/100	99/102	83/88	23/31
Obs. Left/Right	966/273	966/273	730/221	266/105

*** p<0.01, ** p<0.05, * p<0.1; Standard errors in parentheses
"Obs Left/Right" indicates the total number of observations to the left/right of the threshold.
"Eff. Obs. Left/Right" refers to the number of observations effectively used in estimations (inside the bandwidth).

We repeat these exercises for the 75% threshold (lower panel in Table 1). As previously discussed, we use a fuzzy approach in this case. Despite differences in magnitudes, the message is broadly the same. In none of the specifications, we find the relevant coefficient to be statistically significant.[2] In other words, we do not find evidence of discontinuity in either cutoff employed by RT.

These results are line with Eliashberg and Shugan (1997), Reinstein and Snyder (2005), and Thrane (2018), which conclude that critics have little influence, especially for wide releases. We show that this also holds for an online aggregator such as RT.

[2] We consider other kernels to weigh observations and non-linearities in equation (1) – through polynomials up to order 4. Qualitatively, our key results remain the same.

A possible interpretation is that the public has access to other sources of information about wide releases, which feature high profile casts and larger advertisement budgets. This could limit the effect of critics' opinions on moviegoers' choices. It remains an open question whether RT has a different impact on smaller-scale productions. However, given our motivation and data availability issues associated with limited releases, we decided to focus on wide releases in this paper.

4. Conclusion

Can bad reviews lead movies to tank at the box office? Recently, this reasoning has been extended to movie review aggregators. Here we study Rotten Tomatoes (RT), a popular aggregator that is often blamed by the movie industry for the failure of certain productions. RT's classification of movies into discrete categories allows us to use regression discontinuity techniques, which addresses omitted variable issues associated with unobservable quality or commercial appeal. Using a sample of more than 1,200 films widely released in the U.S. between 1999 and 2019, we find no evidence that RT scores drive box office performance.

5. References

Basuroy, S., Abraham Ravid, S., Gretz, R.T. et al. 2020. Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues. J Cult Econ 44, 57–96.

Brown, A.L., Colin F.C., Lovallo, D. 2012. To Review or Not to Review? Limited Strategic Thinking at the Movie Box Office. *AEJ: Microeconomics*, 4(2):1-26.

Calonico, S., Cattaneo, M., Farrell, M., Titiunik, R., 2016. rdrobust: Software for regression discontinuity design. Stata J. 1–30.

Cattaneo, M.; Idrobo, N.; Titiunik, R. 2020. A Practical Introduction to Regression Discontinuity Designs: Foundations (Elements in Quantitative and Computational Methods for the Social Sciences). Cambridge University Press. doi:10.1017/9781108684606

Eliashberg, J.; Shugan, S. (1997). Film Critics: Influencers or Predictors? Journal of Marketing,

61(2), 68-78.
Ginsburgh, V. 2003. "Awards, Success and Aesthetic Quality in the Arts." Journal of Economic Perspectives, 17(2):99-111.
Reinstein, D. A., Snyder, C. M. 2005. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. Journal of Industrial Economics, 53(1): 27–51.
Thrane, C. (2018). Do expert reviews affect the decision to see motion pictures in movie theatres? An experimental approach. Applied Economics, 50(28):3066-3075.