RedFame

Why do Movie Studios Produce R-rated Films?

Brian Goff[1], Dennis Wilson [1]& David Zimmer[1]

[1]Department of Economics, Gordon Ford College of Business, Western Kentucky University, United States.

Correspondence: David Zimmer, Department of Economics, Gordon Ford College of Business, Western Kentucky University, United States.

Abstract

R-rated films are correlated with lower box office revenues. Using sex and nudity content as an instrument to predict R-ratings, we show that R-ratings themselves are not the cause of lower revenues. Instead, R-rated films contain traits that lower revenues and the ratings act as a signal of those traits. Why do film producers include those traits and seek R-ratings? We show that R-ratings alter the shape of the distribution of critical reviews, shifting the distribution to the right. Moreover, even after controlling for other influences, R-ratings improve critical reviews, suggesting that producers not only seek box office revenues, but also critical acclaim.

Keywords: monopolistic quality choice, causal impact, quantile treatment effects

1. Introduction

In the summer of 2012, Universal Pictures released a comedy film in which a child's teddy bear comes to life to become the child's best friend. Although numerous Hollywood films have revolved around similar plots of childhood toys coming to life, this particular movie, entitled *Ted*, was different. It was rated R. In the months leading up to its release, some commentators in the popular press considered *Ted* a financial risk, given that rated R films typically underperform their non-rated R counterparts. However, *Ted* was a huge financial success, eventually becoming the globally highest-grossing rated R comedy of all tim

Despite *Ted*'s success, could the film have achieved even greater successful had it not been rated R? On one hand, maybe; after all, most movie theatres do not permit moviegoers younger than 17 into rated R films without parental accompaniment, a restriction that would seem to reduce the potential audience for a movie about a teddy bear. Perhaps a PG or PG-13 rating could have expanded its audience. But on the other hand, much of *Ted*'s comedic appeal involved the seemingly-innocent childhood toy cracking filthy jokes, consuming a variety of illegal drugs, and engaging in and referencing promiscuous sex. Take away the bear's rated R behavior, and it's not obvious how funny he would be.

This paper explores why film studios produce rated R movies. In doing so, this paper draws inspiration from models of monopolistic quality choice, in which firms face potential tradeoffs between quality and profit. In the seminal work on this topic, Spence (1975) develops a model in which firms endogenously set aspects of quality, with such choices potentially leading to market failure. Aghion, Bloom, Blundell, Griffith, and Howitt (2005) explore similar topics. Our paper explores an empirical application of monopolistic quality choice models, with emphasis on the film industry in which a movie's financial success depends, in part, on "quality," which is subjectively determined, but might be signaled by way of the film's rating.

Determining the impact of an R-rating on box office performance is complicated, because a film's rating is not an exogenously-bestowed label. Instead, a rating is an endogenously chosen feature of the film itself, and one that is (approximately) known and targeted by movie studios before beginning production. In this paper, we address two related questions. First, although R-rated films earn less at the box office compared to their non-rated R counterparts, what is the *causal* effect of an R-rating on a film's box office performance? To identify that causal effect, we employ what we believe is a novel identification approach. Specifically, we posit that a film's sexual and/or nudity content significantly influences a film's probability of being rated R, but does not directly impact a film's box office performance. Using sex/nudity as an instrument, we show that a film's R-rating does not causally reduce its box office performance. Rather, the lower performance of rated R movies seems to stem from the fact that rated R films possess specific traits that are difficult to measure quantitatively, but are easy for moviegoers to detect. And it is those

difficult-to-measure traits, rather than the R-rating itself, that result in lower box office performance. We offer several robustness checks, some of which rely upon alternative identification strategies. Those robustness checks appear to support this conclusion.

This finding segues to our second question: If rated R films perform worse at the box office, why do film studios produce so many rated R films? Our answer relies on the idea that market revenues represent only part of the motivation for film making. Movie producers, directors, and actors also seek critical acclaim, whether from professional movie critics or from a subset of the public. The nature of the Motion Picture Association of America's (MPAA) rating system places greater restrictions on artistic expression for films rated G, PG, and PG-13 than for those rated R. Such freedom of artistic expression potentially appeals to those who judge films on artistic merit. To that end, we provide evidence that R-ratings positively correlate with critical acclaim, even after controlling for other film attributes. These results suggest that movie makers may use an R-rating as a signal to professional critics.

2. Background

Determinants of box office revenues have received considerable investigation within economics and marketing. DeVany and Walls provide several contributions (1996, 1997, 1999, 2002, 2004), along with Eliashberg, Elberse, and Leenders (2006), Liu (2006), Elliott and Simmons (2008), Weinstein (1998), Swami, Eliashberg, and Weinberg (1999), and Holbrook (1999).[1] One well-established observation from these studies is that motion picture revenues display fat-tailed distributions. Explanations of the fat tails have included inter-consumer informational effects and non-linearities as in DeVany and Walls (1996), along with the dynamic evolution of film enjoyment for a given consumer as in Eliashberg and Sawhney (1994).

A second observation from extant research is that R-rated films earn less at the box office. DeVany and Walls (2002) show that revenues for R-rated films are lower, but those revenues as a ratio of film costs are about the same. Explanations for this observation are not as widespread. DeVany and Walls attribute it to the fact that R-rated films display even greater uncertainty and risk than non-rated R films, but that certain high-grossing R-rated films generate exceptional prestige, creating an incentive for influential stars to lobby for their production. Their empirical work and explanation raise a more fundamental empirical question: Are R-ratings themselves the source of lower revenues or do R-ratings merely indicate the presence of other factors? That question has not been directly addressed in the literature, yet it is a critical issue in sorting out the role and explanation of R-ratings.

We contend that the production of R-rated movies can be, in part, a revealing of characteristics of the actor's, director's, producer's, and studio's utility functions. The importance of financial success notwithstanding, each agent derives utility from the artistic impression of their craft, even if such artistic success may come at the detriment of potential profitability. Among mainstream movie ratings, the R-rating provides movie producers with the most artistic leeway. Such artistic freedom strays into characteristics that may be difficult to quantify but are nonetheless important to movie creators: tone, style of humor, plot development, character expansion, political themes, artistic "look," and others. However, how far movie makers are willing to push those artistic freedoms does not completely overlook potential monetary gains. For example, because theaters are reluctant to provide prime screening times to movies with NC-17 ratings, several artistically noteworthy movies have been edited to avoid such a rating. Quentin Tarantino removed scenes from "Pulp Fiction" and "Kill Bill" to avoid an NC-17 rating, as did Spike Lee from "Summer of Sam."

Such artistic elements are generally appreciated by the Academy of Motion Pictures and published movie critics. Thus, a movie like "The Artist" wins five Academy Awards, including the Oscar for Best Picture in 2012, while generating less than 1/10th the box office revenues of "The Avengers" and "The Dark Knight Rises" in the same year. The monetary value to movie makers of such critical acclaim and recognition for artistic work is not inconsequential, and our results suggest that producers may actively seek such critical acclaim by pursuing such projects.

3. Data

We collected data on all U.S.-released films over 2000-2014 which made in excess of $1 million in gross box office revenue, and for which relevant data were available.[2] We use the $1 million threshold because many films below that threshold do not receive theatrical releases, and therefore do not subject themselves to the MPAA ratings process. Furthermore, data for films below that level are often unavailable or incomplete, indicating that those films have such limited releases that both consumers and reviewers are often unaware of their existence. We consider only films with MPAA-assigned ratings. We do not consider films with NC-17 ratings, as such films typically do not have widespread

---

[1] Eliashberg (2005) provides an extensive background to the industry and literature about it.

[2] The 2014 data are for part of the year.   Box office, rating, screenplay source, genre category, and release date are from Internet Movie Database available at http://www.imdb.com/ .

releases. Our final sample consists of 1,784 films.

Our outcome variable of interest is U.S. box office revenue. Of course, films also earn revenues from overseas releases and DVD/merchandise sales, but those outlets do not use or emphasize MPAA ratings in the same way as U.S. theatres. Therefore, we do not consider revenue from those sources. The first row of Table 1 shows mean revenue for rated R films and for G, PG, and PG-13 films (hereafter collectively called "non-rated R.") As reported in previous research, non-rated R films perform far better at the box office.

Table 1. Sample Means for 1,784 Films Released 2000-2014

	Rated R N = 736	Not Rated R N = 1,048	Significantly Different?
U.S. box office revenue	36,800,000	70,00,000	**
Budget	29,000,000	57,500,000	**
Political content	0.05	0.02	**
Based on book	0.19	0.27	**
Original screenplay	0.60	0.48	**
Horror	0.12	0.04	**
Romantic	0.07	0.12	**
Action	0.19	0.23	**
Animated	0.01	0.10	**
Month of release	12 dummies	12 dummies	

* $p < .10$ ** $p < .05$

However, we cannot conclude from Table 1 that an R-rating *causes* lower box office performance, because as shown in the remainder of the Table, other movie characteristics also show significant variation with respect to R-ratings. Most noteworthy, rated R films have smaller budgets. Aside from budget, rated R films differ from their non-rated R counterparts across other dimensions. Specifically, R-rated films include a higher proportion of original screenplays and horror films, but smaller proportions of romance, action, animated movies. Also, R-rated films are more likely to include political content, defined as films that receive nominations from the Political Film Society.[3]

The fact that measurable attributes differ with respect to ratings raises the possibility that R-rated films also display other unique characteristics that are not easily quantifiable. The presence of such unobserved attributes clouds the relationship between an R rating and box office performance. The following section attempts to assemble an econometric model capable of addressing those unobserved traits.

4. Methods

Our estimations are based on the model

$$\ln \text{box}_i = \gamma Rated\ R_i + \mathbf{X}_i'\boldsymbol{\beta} + \mathbf{U}_i'\boldsymbol{\delta} + \varepsilon_i$$

where $\ln \text{box}_i$ denotes the natural logarithm of U.S. box office revenue for film i, and ε_i represents an error term. The variable $Rated\ R_i$ is a dichotomous indicator equaling 1 if film i is rated R, and 0 otherwise. Our main interest lies in the parameter γ, which captures the extent, if any, to which an R-rating affects box office revenues. We partition control variables into two separate vectors, with the vector $\mathbf{X}_i'$ consisting of film traits that we observe, and the vector $\mathbf{U}_i'$ consisting of film traits that we cannot observe. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ include coefficients attached to those traits.

We seek to interpret the coefficient γ as a causal parameter, but such an interpretation becomes difficult due to the presence of $\mathbf{U}_i'$. Because we do not observe components of that vector, and therefore cannot include them during estimation, those unobserved traits become absorbed into the composite error term $(\mathbf{U}_i'\boldsymbol{\delta} + \varepsilon_i)$. Interpretation of γ as a causal parameter requires that $E(\mathbf{U}_i'\boldsymbol{\delta} + \varepsilon_i \mid \text{Rated R}_i) = 0$. This condition is violated if, for example, rated R films tend to have more serious or artistic tones.

One path to consistent estimation of γ requires an "instrument" that significantly impacts the likelihood that a film receives an R-rating, but at the same time does not belong in the regression equation. For such an instrument, we use a dichotomous indicator of whether the film contains sexual content and/or nudity (hereafter shortened to "sex/nudity").[4] According to the MPAA's description of its ratings system, the presence of sex/nudity in a film, depending on its quantity and context, might warrant an R-rating.[5] Furthermore, the MPAA has been frequently accused of assigning

---

[3] These data are available at http://www.polfilms.com/previous.html.

[4] Sex/nudity content was categorized by Yahoo! Movies.

[5] See http://www.mpaa.org/ratings

R-ratings to films that feature modest amounts of sex/nudity, while taking a much softer approach toward violence.[6] We avoid wading into that debate. Rather, we exploit the supposition that sex/nudity significantly influences the likelihood that a film receives an R-rating. At the same time, it seems plausible that, after controlling for a film's rating, as well as other film attributes, the presence (or lack) of sex/nudity should not significantly alter a film's box office performance. The following subsections investigate the validity of these suppositions.

4.1 First-Stage Instrument Validity

For statistical validity, the instrument sex/nudity must not be "weak," which means that its impact on a film's rating must be quantitatively large and statistically significant. This is especially important because, to preview our result, we find that the coefficient of $Rated\ R_i$ becomes statistically insignificant after instrumenting. We choose to interpret that finding as evidence that an R-rating does not causally impact a film's box office performance, but such statistical insignificance also can be an artifact of a weak instrument. Therefore, we wish to rule out that possibility.

To demonstrate instrument strength, we regress $Rated\ R_i$ on all explanatory variables, including sex/nudity, with results appearing in Table 2. The coefficient of $Rated\ R_i$ is 0.191. The interpretation is that the presence of sex/nudity correlates with a 19.1 percentage point increase in the probability that a film receives an R-rating. Given that 41 percent of films in our data have R-ratings, this 19.1 percentage point increase implies that sex/nudity increases the probability of an R-rating by almost 47 percent relative to the sample mean. Furthermore, the coefficient of $Rated\ R_i$ is statistically significantly different from zero. Its associated F-statistic is 69.3, which exceeds the commonly used threshold of 10 to denote instrument strength (Staiger & Stock, 1997). The F-statistic associated with the more conservative Stock and Yogo (2005) test is 70.6, which exceeds the critical value of 16.38, assuming a willingness to tolerate distortion on a 5 percent Wald test of no more than 10 percent. (Note, the Stock and Yogo result should be interpreted with caution, as it assumes homoscedastic errors, whereas we detect heteroscedasticity in our data.) All together, we interpret these findings as evidence that sex/nudity exerts a quantitatively large and statistically significant impact on the likelihood of R-ratings. In results presented below, we employ alternative instruments, which even more overwhelmingly reject the null hypothesis of weak instruments. Results from those models are nearly identical to our baseline findings, which we interpret as offering further evidence of instrument strength.

Table 2. OLS Regression Dependent Variable: Rated R

	Coeff	St Err
Sex/nudity	**0.191****	**0.023**
Log budget	−0.099**	0.010
Political content	0.197**	0.058
Based on book	−0.022	0.031
Original screenplay	0.061**	0.027
Horror	0.236**	0.041
Romantic	−0.166**	0.036
Action	0.088**	0.028
Animated	−0.228**	0.047
Constant	2.005**	0.178
Month of release dummies?	yes	

* p < .10 ** p < .05

4.2 Second-Stage Instrument Validity

For sex/nudity to serve as a valid instrument, it must be excludable from the regression equation, meaning that it must not correlate with the composite error term $(\mathbf{U}_i'\boldsymbol{\delta} + \varepsilon_i)$. To investigate the excludability assumption, we estimate the regression equation by simple OLS, and then regress the predicted residuals from that regression on sex/nudity (and a constant). Results from that regression, presented in Table 3, reveal virtually zero correlation between sex/nudity and the residuals from the regression equation. Consequently, it appears that sex/nudity does not correlate with box office performance, other than indirectly via a film's rating.

Table 3. OLS Regression Dependent Variable: Second-Stage Residuals

	Coeff	St Err
Sex/nudity	**0.028**	**0.053**
Constant	−0.016	0.040

---

[6] See "Roger Ebert Thinks the MPAA's Ratings are Useless."

http://www.joblo.com/movie-news/roger-ebert-thinks-the-mpaas-ratings-are-useless

There does not exist a formal test for second-stage instrument validity in a just-identified model, so we interpret Table 3 as providing informal support for our identification assumption. However, in results presented later in the paper, we employ additional instruments, such that the models become overidentified, thus permitting formal Hansen (1982) tests of overidentification. Results from those tests appear to support second-stage instrument validity.

4.3 LATE Interpretation of IV Estimate

Using the sex/nudity instrument, we seek to estimate the causal impact of an R-rating on box office revenue. However, that causal impact is likely to vary across films. Consequently, we interpret the coefficient γ in the regression equation as the average impact for films that cross the threshold into rated-R status *due to* their sex/nudity content. This is the local average treatment effect (LATE) interpretation highlighted by Imbens and Angrist (1994).

Additionally, in results presented later, we employ alternative instruments, which necessarily change the interpretation of the LATE estimator. Nonetheless, estimates based on those alternative instruments are similar to our baseline findings using sex/nudity. Consequently, we have some confidence that our main findings extrapolate to a large group of films, rather than just those that receive R-ratings due to their sex/nudity content.

5. Results

Table 4 presents estimates for regressions of log U.S. box office revenues. The left panel shows OLS estimates, which do not correct for potential endogeneity of $Rated\ R_i$. The right panel shows IV estimates, which instrument for $Rated\ R_i$ using an indicator for whether the film contains sex/nudity. For the most part, coefficients of the control variables corroborate *a priori* expectations, both in terms of sign and magnitude. Specifically, films with larger budgets tend to perform better at the box office. Moreover, action and animated movies perform better at the box office, as do horror films. (This latter finding might explain the seemingly never-ending output of critically-derided slasher flicks.)

Table 4. OLS and IV Regressions Dependent Variable: Log U.S. Box Office Revenue Instrument in IV Model = Does Film Contain Sex and/or Nudity?

	OLS		IV	
	Coeff	St Err	Coeff	St Err
Rated R	**−0.268**[**]	**0.062**	**−0.098**	**0.306**
Log budget	0.713[**]	0.032	0.732[**]	0.047
Political content	−0.145	0.162	−0.175	0.166
Based on book	−0.038	0.077	−0.035	0.078
Original screenplay	0.040	0.073	0.028	0.074
Horror	0.661[**]	0.099	0.624[**]	0.121
Romantic	0.021	0.089	0.042	0.095
Action	0.135[**]	0.064	0.122[*]	0.067
Animated	0.163[*]	0.097	0.217	0.135
Constant	4.900[**]	0.589	4.514[**]	0.931
Month of release dummies?	yes		yes	
R-squared	0.42		0.42	

* $p < .10$ ** $p < .05$

Our main interest lies in the coefficients of $Rated\ R_i$, which appear near the top of Table 4. The OLS coefficient reveals that films with R ratings earn approximately 27 percent less than non R-rated movies. That estimate corroborates the simple sample averages reported in Table 1. However, as argued above, a film's rating is not an exogenously-bestowed characteristic, but rather an intrinsic indicator of a film's tone and content. Director, actors, and producers choose to work on films with approximate ratings, and attendant box office potential, already in mind. That is, an R rating likely is endogenous with respect to box office performance.

The IV estimate of the coefficient of $Rated\ R_i$ shrinks in magnitude relative to its OLS counterpart and loses statistical significance. The IV estimate implies that having an R-rating does not causally hinder a film's box office performance. Rather, rated R films possess traits that correlate with the rating, and it is those traits, rather than the rating itself, that repel some moviegoers. We explore this finding in more detail in Section 7.

6. Alternative Specifications

In this section, we investigate potential threats to the validity of the estimates presented in the previous section.

6.1 Alternative Instruments

We prefer sex/nudity as an instrument both on an intuitive level, and because it is easy to objectively determine whether films contain sex/nudity. However, we acknowledge the existence of (at least) two other potential instruments: (1)

whether a film contains violence and (2) whether a film contains objectionable language. Both of these indicators obviously impact a film's likelihood of receiving an R-rating, and likewise, both of these indicators are unlikely to impact a film's box office performance, other than indirectly through its rating.

Our concern with these two potential instruments, and the main reason we do not include them in our baseline models, centers on the difficulty of objectively determining whether a film contains such material. What qualifies as violence? Must someone be harmed physically? Must weapons be involved? Does pushing and shoving, if performed in a threatening manner, count as violence? What about violence conducted in a comedic manner? And what about mere threats of violence? (Curiously, the groundbreaking 1978 film *Halloween*, which inspired a generation of extremely violent slasher films, did not itself contain extensive violence. But it did contain suspenseful scenes in which characters were scared for their physical safety. Are such scenes "violent?") Similar problems exist in defining objectionable language. What words count as objectionable, and in what quantity are they deemed objectionable? Does the tone in which those words are spoken matter? Does threatening or suggestion language count as objectionable, even if no obvious flag words are used?

In sum, we fear that too much latitude exists in how one defines "violence" and "language," such that results obtained from models using those instruments should be viewed with a degree of skepticism. Nonetheless, for the purpose of exploring robustness of our baseline findings, we attempt, as objectively as we can, to assign each film in our data indicators of whether they contain violence or objectionable language. A movie is listed as containing violence and/or objectionable language if described as such by the brief ratings defense provided in support of the movie's rating by the Classification and Rating Administration of the MPAA.   These definitions are somewhat liberal, resulting in 52 percent of the films being labeled as "violent," and 69 percent flagged as containing objectionable language.

The advantage of now having three instruments (sex/nudity, violence, and language) is that our model is overidentified. Therefore, we can re-estimate the models using different combinations of instruments, while also calculating Hansen (1982) tests of instrument validity. Results from those alternative models appear in Table 5. The coefficient of $Rated\ R_i$ remains similar to our baseline estimates, and does not achieve statistical significance for any combination of instruments. The other panels of the table confirm instrument validity. The F-tests, as well as the more conservative Stock and Yogo tests, both allow overwhelming rejection of the null hypothesis of weak instruments, even more so than in our baseline model. And the Hansen tests do not allow rejection of the null hypothesis that the instruments impact box office performance, which is not surprising in light of the overall insignificance of $Rated\ R_i$. In total, we view these alternative models as offering support to our baseline findings.

Table 5. Robustness Checks Using Different Identification Strategies Impact of Rated R on Log U.S. Box Office Revenues

	Coeff	St Err	F-test for first stage joint significance of of instruments	Stock and Yogo F-test for first stage joint significant of instruments	Hansen chi-square test for second stage instrument exogeneity
IV models (instruments)					
sex/nudity, violence	−0.166	0.244	64.7	59.8	0.14
sex/nudity, language	−0.075	0.211	99.7	88.8	0.01
violence, language	−0.119	0.206	96.2	87.0	0.23
sex/nudity, violence, language	−0.118	0.189	84.9	74.9	0.23
Nearest-neighbor match model	−0.092	0.097	--	--	--

6.2 Matching Estimator

Instrument validity checks notwithstanding, it is instructive to consider an alternative identification strategy that does not rely upon instruments. Therefore, using a probit estimator and all of the explanatory variables, each film is assigned a predicted probability of receiving an R-rating (i.e., a propensity score). Then, each actual Rated R film is matched (with replacement) to a non-Rated R film with the nearest propensity score.

After matching, the difference in mean log box office revenues between the two groups provides an estimator of the impact of R-ratings. The associated standard errors are calculated by bootstrap. (Various alternative matching methods, such as radius methods with different calipers, produced similar findings.)

Identification of the matching estimator assumes that films with close propensity scores of R-ratings share identical unmeasured characteristics. The matching estimator provides the average treatment effect on the treated, as opposed to the local average treatment effect produced by the IV approach. Nonetheless, as reported in the bottom row of Table 5, the matching estimator produces a similar conclusion to that obtained from the IV model.

6.3 Quantile Regressions

In this subsection, we investigate whether basic linear regression methods are appropriate in light of the highly-skewed distributional shape of box office revenues. A common procedure, and the one employed in this paper, is to calculate the natural logarithm of the outcome measure, and use the logged version as the dependent variable of interest. Yet, previous research on box office revenue – beginning with DeVany and Walls (1996) – suggests that the degree of skewness and kurtosis inherent in box office revenues renders even log revenues severely non-symmetric. DeVany and Walls show that week-by-week revenues during a film's run follow Bose-Einstein dynamics, with final accumulated revenues converging to a stable Pareto distribution.

We avoid employing maximum likelihood estimations using those alternative distributions, because methods for addressing endogeneity in those models remain in their infancy. Instead, we explore distributional concerns of box office revenues using quantile regression methods. Although our main concern is in exploring the robustness of our baseline results, quantile regression methods offer additional interesting insights in allowing us to determine whether the impact of an R-rating varies over the distribution of box office revenues (Koenker & Bassett, 1978).

Table 6 presents quantile regression estimates for several quintiles of log U.S. box office revenues. The middle panel shows estimates for a median regression (i.e., .50 quantile). The coefficient of $Rated\ R_i$ from the median regression is similar in magnitude to the corresponding coefficient from the OLS model (Table 4). Estimates from the other quintiles show that the extent to which an R-rating correlates with smaller box office revenues is somewhat more pronounced for lower-earning films. Specifically, an R-rating for films that earn near the 0.10 quantile correlates with an approximate 33 percent reduction in box office performance. On the other hand, an R-rating for films near the 0.90 quantile suffer only a 20 percent drop in box office performance. Some other explanatory variables also show variation across quintiles. For example, the boost enjoyed by horror movies is most pronounced among the lowest-earning films, although the highest-earning films also appear to enjoy a horror-film premium.

Table 6. Quantile Regressions Dependent Variable: Log U.S. Box Office Revenue

	.10 quantile		.25 quantile		.50 quantile		.75 quantile		.90 quantile	
	Coeff	St Err	Coeff	St Err	Coeff	St Err	Coeff	St Err	Coeff	St Err
Rated R	-0.334^{**}	0.120	-0.240^{**}	0.087	-0.225^{**}	0.064	-0.237^{**}	0.058	-0.202^{**}	0.064
Log budget	1.056^{**}	0.052	0.919^{**}	0.038	0.712^{**}	0.028	0.597^{**}	0.025	0.515^{**}	0.028
Political content	-0.043	0.299	-0.174	0.216	-0.199	0.159	-0.289^{**}	0.144	-0.005	0.160
Based on book	0.017	0.160	-0.021	0.116	-0.150^{*}	0.085	-0.056	0.077	-0.008	0.086
Original screenplay	0.085	0.138	0.091	0.100	-0.081	0.073	-0.007	0.067	0.051	0.074
Horror	0.864^{**}	0.211	0.720^{**}	0.153	0.419^{**}	0.112	0.283^{**}	0.102	0.500^{**}	0.113
Romantic	0.234	0.185	0.068	0.134	0.080	0.098	-0.018	0.089	-0.182^{*}	0.099
Action	0.145	0.145	0.124	0.105	0.093	0.077	0.115	0.070	-0.026	0.078
Animated	0.361	0.237	0.138	0.171	0.196	0.126	0.118	0.114	0.023	0.127
Constant	-2.734^{**}	0.938	0.807	0.769	5.188^{**}	0.499	7.810^{**}	0.454	9.606^{**}	0.502
Month of release dummies?	yes		yes		Yes		yes		Yes	

* $p < .10$ ** $p < .05$

To account for endogeneity of an R-rating, we estimate IV quantile treatment effects models (Abadie, Angrist, & Imbens, 2002) using sex/nudity as the instrument. Those estimates appear in Table 7. Similar to our baseline findings, the coefficient of $Rated\ R_i$ loses significance at all quantiles. In sum, quantile regression estimates, while shedding light on certain details that remain hidden in basic linear regression models, do not point to any obvious specification errors in our baseline models.[7]

Table 7. IV Quantile Treatment Effects Regressions Dependent Variable: Log U.S. Box Office Revenue Instrument = Does Film Contain Sex and/or Nudity?

	.10 quantile		.25 quantile		.50 quantile		.75 quantile		.90 quantile	
	Coeff	St Err	Coeff	St Err	Coeff	St Err	Coeff	St Err	Coeff	St Err
Rated R	-0.001	0.505	-0.070	0.396	-0.074	0.385	-0.067	0.340	-0.048	0.348
Explanatory variables	yes		yes		yes		yes		yes	
Month of release dummies?	yes		yes		yes		yes		yes	

* $p < .10$ ** $p < .05$

---

[7] To check robustness of the standard errors, all quantile models were re-estimated using a bootstrap approach. Standard errors from those checks were nearly identical to those appearing in the tables.

7. Why Do Studios Produce R-Rated Films?

The estimates in the previous sections suggest that R-rated films perform worse at the box office, but that that effect does not appear to be causal. Rather, R-rated films possess traits, aside from those we are able to control for, and it is those traits that result in lower box office revenue. Such traits might reflect a film's general atmosphere, such as tone, mood, artiness, and degree of "grit." Or those traits could capture a film's technical visual style. For example, famed directors Stanley Kubrick and Quentin Tarantino are both renown for films that achieve specific "looks." Those traits are not easy to quantify and are therefore difficult to control for in regression settings. Prospective moviegoers, however, can detect those traits by viewing movie previews, reading critical reviews, or communicating with others moviegoers.

Whatever those traits might be, their presence in a movie results in lower box office performance.[8] That result raises an obvious question: Why do movie studios produce such films? After all, film studios do show some aversion to sacrificing box office success for the sake of artistic expression. Most notably, studios seek to avoid the rating that offers the most artistic license, NC-17, and will edit films to avoid that rating, as with the aforementioned "Pulp Fiction" and "Summer of Sam." But why do studios produce movies with *any* traits known to correlate with lower box office performance?

We posit that studios produce films with potentially revenue-lessening traits in order to achieve critical acclaim, whether among professional critics, among artistic peers, or among a narrow, highly discriminating section of the market. While this tradeoff between commercial success and critical or artistic acclaim deviates from a simple economic model of motivation, it is not at odds with a broader, Spence-type model of monopolistic quality choice (Spence, 1975; Aghion et al., 2005). That literature illustrates that firms often face tradeoffs between quality and revenue. For example, internet search engines must find the optimal mix of unpaid search results, which users view as "quality," and revenue-generating ads, which might detract from perceived quality (White, 2013). For the movie industry, although financial success clearly matters to Hollywood executives, films studios likely have utility functions that depend on (some) nonfinancial elements, such as reputation. Critical or artistic acclaim as a preference in artistic production can be observed outside of film production in areas such music, painting, and sculpture. Within "alternative" music markets and some other niches, for example, widespread market popularity is sometimes viewed negatively by artists or fans. For them, "pop" is a pejorative term. In the music industry, publications such as Rolling Stone grew up by promoting artists outside of the mainstream.

In addition to purely preference-driven motivations in seeking critical acclaim, there may be long run interactions between critical acclaim, market revenue, and the skills of directors and other participants in movie production. With artistic skill distributed differently across individuals, some directors may hold a comparative advantage in the production of "artistic" films that complement preferences for them. Also, high critical acclaim, especially if leading to award nominations and awards, may feed back into revenues. While the producer and director, in choosing an R-rating, may willingly forego the chance for "blockbuster" size revenue, they can make a relatively inexpensive movie that may achieve substantial profitability at much lower revenue levels, especially if those revenue levels are boosted by critical acclaim and award nominations. For instance, directors Joel and Ethan Coen are well-known for making movies with R-ratings that have relatively low budgets but receive critical acclaim, cult-like followings, and large profit margins. Their 2007 "No Country for Old Men," which won an Academy Award for Best Picture, cost only $25 million to make but grossed nearly $75 million in the U.S. Likewise, their 1996 film "Fargo" also won the Best Picture award on a budget of only $7 million. Furthermore, directors and actors in critically acclaimed films often boost their careers and future earnings. For example, "Pulp Fiction" is widely credited with reviving John Travolta's career.[9]

To investigate the relationship between an R-rating and critical reception, for each film in our estimation sample, we collected its critical review from the Rotten Tomatoes "Tomatometer," which aggregates reviews of "top critics" into a number on the 0-100 scale. To qualify as "top," Rotten Tomatoes requires that a critic must be published in an outlet ranking in the top 10 percent of circulation, employed as a film critic at a national broadcast outlet for at least five years, of employed at an editorial-based website with at least 1.5 million monthly unique visitors for at least three years. We did not include this measure of a film's critical response in our box office regressions, in part due to the endogenous nature of critical reviews. Previous research indicates that critical reviews and box office success both depend on unmeasured film traits as well as unobserved differences between critics' and consumers' quality perceptions.

---

[8] DeVany and Walls (2002) show that revenues for R-rated films are lower but that revenues as a ratio of costs are about the same.

[9] See, for example, http://www.rollingstone.com/movies/reviews/pulp-fiction-19941014. The film is also credited with reviving the career of Bruce Willis.

Consequently, standard regression setups that condition film success on critical reviews become difficult to interpret.[10] Rather, we use the critical review measure to test our hypothesis that R-rated movies garner higher critical respect.
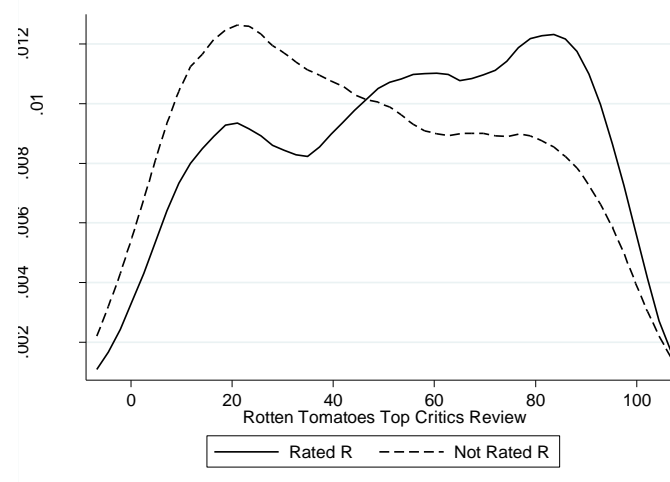

Figure 1. Empirical Distribution of Critical Reviews

Figure 1 shows empirical distributions of critical reviews. The distribution of reviews for non-rated R films appears to be leftward skewed, with a mode review around 20. Rated R films, by contrast, are skewed to the right, with a mode review around 80. Table 8 presents similar evidence by regressing critical reviews on movie traits. Regardless of whether we include other controls, a rated R movie enjoys a boost in critical response of approximately 8-9 points on the 0-100 scale. The median review across all films in our data is 50, so the estimated boost of 8-9 points amounts to an approximate 16-18 percent improvement in critical response, relative to the sample median.

Table 8. Regression of Critical Review on Movie Traits

	Coeff	St Err	Coeff	St Err
Rated R	**8.33**[**]	**1.36**	**8.62**[**]	**1.38**
Log budget			-4.83[**]	0.60
Political content			10.72[*]	3.42
Based on book			7.00[**]	1.83
Original screenplay			-1.42	1.58
Horror			-17.38[**]	2.42
Romantic			-6.09[**]	2.12
Action			-0.60	1.66
Animated			16.61[**]	2.71
Constant	45.97[**]	0.87	137.22[**]	10.75

$R^2 = 0.02$

* $p < .10$ ** $p < .05$

As for other coefficients in Table 8, critics appear to dock big budgets films, as well as films that fall in the popular horror and romantic genres. On the other hand, critics react warmly to animated features and movies based on books. Moreover, critics appear to favor films that dare to include some form of political content. Taken as a whole, the estimates in Table 8 suggest that critics prefer films that do not aim for mainstream success.

The results of this section suggest that, despite lower revenues earned by rated R films, studios release rated R projects because such films are more likely to garner critical praise. We do not argue that an R-rating causally improves a film's critical response. Rather, films with R-ratings, and whatever traits might associate with those R-ratings, tend to appeal to critics.   Besides the utility gained by artists from critical acclaim, high reviews might have impacts on the long run financial rewards and future critical acclaim available to producers, directors, and actors who receive such acclaim. It may generate a greater likelihood of finding future projects, higher quality future projects, Oscar nominations, expanded social networks, and other non-monetary benefits.

---

[10] Eliashberg and Shugan (1997); Reinstein and Snyder (2005); Holbrook and Addis (2008); Hennig-Thurau, Marchand, and Hiller (2012)

8. Conclusion

Our results indicate that focusing on R-ratings in themselves stops short of a richer understanding of movie markets. Rather, ratings act as indicators of other traits. Ratings are also endogenous to film production choices that seem to tradeoff commercial success and critical acclaim. An extension of this line or reasoning raises the questions as to the extent that critics themselves are part of the movie market. In many respects, movie makers and critics are sharing a market. By offering high acclaim for films with low commercial appeal, critics may by selling their services to the "cinephile" part of the movie market.

References

Abadie, A., Angrist, J., & Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, *70*, 91-117. http://dx.doi.org/10.1111/1468-0262.00270

Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: an inverted-U relationship. *Quarterly Journal of Economics*, *120*, 701-728.

DeVany, A., & Walls, W. (2004). Motion picture profit, the stable paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics and Control*, *28*, 1035-1057. http://dx.doi.org/10.1016/S0165-1889(03)00065-4

DeVany, A., & Walls, W. (2002). Does Hollywood make too many R-rated movies? risk, stochastic dominance, and the illusion of expectation. *Journal of Business*, *75*, 425-451.

DeVany, A., & Walls, W. (1999). Uncertainty in the movie industry: does star power reduce the terror of the box office? *Journal of Cultural Economics*, *23*, 285-318. http://dx.doi.org/10.1023/A:1007608125988

DeVany, A., & Walls, W. (1997). The market for motion pictures: rank, revenue, and survival. *Economic Inquiry*, *35*, 783-797. http://dx.doi.org/10.1111/j.1465-7295.1997.tb01964.x

DeVany, A., & Walls, W. (1996). Bose-Einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal*, *106*, 1493-1514. http://dx.doi.org/10.2307/2235197

Eliashberg, J. (2005). The film exhibition business: critical issues, practice, and research. In C. Moul (Ed.), *A Concise Handbook of Movie Industry Economics*, New York, NY: Cambridge University Press, 138-162.

Eliashberg, J., Elberse, A., & Leenders, M. (2006). The motion picture industry: critical issues in practice, current research, and new recent directions. *Marketing Science*, *25*, 638-661.

Eliashberg, J., & Sawhney, M. (1994). Modeling goes to Hollywood: predicting individual differences in movie enjoyment. *Management Science*, *40*, 1151-1173. http://dx.doi.org/10.1287/mnsc.40.9.1151

Eliashberg, J., & Shugan, S. (1997). Film critics: influencers or predictors? *Journal of Marketing*, *61*, 68-78. http://dx.doi.org/10.2307/1251831

Elliott, C., & Simmons, R. (2008). Determinants of UK box office success: the impact of quality signals. *Review of Industrial Organization*, *33*, 93-111. http://dx.doi.org/10.1007/s11151-008-9181-0

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 1029-1054.

Hennis-Thurau, T., Marchand, A., & Hiller, B. (2012). The relationship between reviewer judgments and motion picture success: re-analysis and extension. *Journal of Cultural Economics*, *36*, 249-283. http://dx.doi.org/10.1007/s10824-012-9172-8

Holbrook, M. (1999). Popular appeal versus expert judgements of motion pictures. *Journal of Consumer Research*, *26*, 144-155. http://dx.doi.org/10.1086/209556

Holbrook, M., & Addis, M. (2008). Art versus commerce in the movie industry: a two-path model of motion-picture success. *Journal of Cultural Economics*, *32,* 87-107. http://dx.doi.org/10.1007/s10824-007-9059-2

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467-475. http://dx.doi.org/10.2307/2951620

Koenker, R., & Bassett, G. (1978). Regression quantiles. *Econometrica*, 46, 33-50. http://dx.doi.org/10.2307/1913643

Liu, Y. (2006). Word of mouth for movies: its dynamics and impact on box office revenues. *Journal of Marketing*, *70,* 74-89. http://dx.doi.org/10.1509/jmkg.70.3.74

Reinstein, D., & Snyder, C. (2005). The influence of expert reviews on consumer demand for experience goods: a case study of movie critics. *Journal of Industrial Organization*, *53*, 27-51. http://dx.doi.org/10.1111/j.0022-1821.2005.00244.x

Spence, M. (1975). Monopoly, quality, and regulation. *Bell Journal of Economics*, *6,* 417-429.

http://dx.doi.org/10.2307/3003237

Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65,* 557-586. http://dx.doi.org/10.3386/t0151

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. Andrews, J. Stock, & T. Rothenberg (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, New York, NY: Cambridge University Press., 80-108.

Swami, S., Eliashberg, J., & Weinberg, C. (1999). Silver screener: a modeling approach to movie screens management. *Marketing Science*, 18, 352-372.

Weinstein, M. (1998). Profit-sharing contracts in Hollywood: evolution and analysis. *Journal of Legal Studies*, *27*, 67-112. http://dx.doi.org/10.1086/468014

White, A. (2013). Search engines: left side quality versus right side profits. *International Journal of Industrial Organization*, *31*, 690–701. http://dx.doi.org/10.1016/j.ijindorg.2013.04.003