

The Impact of Originality in a Transitioning Movie Industry

Jacob Graber-Lipperman

Dr. Kent Kimbrough, Faculty Advisor

*Honors thesis submitted in partial fulfillment of the requirements for Graduation with
Distinction in Economics in Trinity College of Duke University*

Duke University
Durham, North Carolina
2019

Acknowledgements

Thank you to Dr. Kimbrough for offering incredible advice throughout the writing and research of my thesis and reviewing several of my drafts over the course of the past academic year. Without your help, I would have struggled mightily to finish this challenging task. You made our Honors Seminar a welcoming place to run our ideas by both you and our classmates, and with your guidance, this course has been one of the most enjoyable I have taken at Duke. I will miss your unparalleled knowledge of Duke Basketball.

I'd like to thank my classmates during both the Fall and Spring semesters of the Economics Honors Seminar for workshopping my research and providing helpful input throughout the process. I'd also like to thank Dr. Adam Rosen for taking time to meet with me and discuss the theory behind my topic as well as several econometric techniques I could utilize when approaching my analysis. Last, I'd like to thank Thomas Howell and Tye Pomerantz for aiding me in the data collection stage when I faced an unexpected roadblock to acquiring the data I needed to complete the fourth test in this paper. Their willingness to volunteer on a moment's notice helped make the completion of my thesis a much smoother endeavor.

One more special thank you goes out to my mom, who pushed me to complete an Honors Thesis instead of coasting through senior year. You always know how to motivate me, and I'm all the wiser for having followed your advice.

Abstract

The thesis explores the increasing success of non-original films distributed through traditional theatrical releases, and asks whether new distributors, such as Netflix, may serve as better platforms for original content. A dataset incorporating the top 100 highest-grossing films at the domestic box office each year from 2000 to 2018, as well as a smaller subset including 82 titles distributed by Netflix, was utilized to investigate these issues. The results confirm non-original content has performed increasingly well over time for theatrical releases, especially within the past four years, while original content has performed poorly, especially during this recent time period. Additionally, the research suggests the stark difference in performance observed for non-original and original content in traditional distribution models may not appear for titles released through the newer streaming platforms. This paper thus hopes to motivate future study into the effect of streaming platforms on consumer purchasing behavior of films as new distribution technology within the movie industry continues to proliferate.

JEL Classification: D1; D10; D19.

Keywords: Film Industry.

I. The Changing Landscape of the American Entertainment Industry

The composition of the American film industry has changed dramatically since the days of *Jaws* and *Star Wars*, the first Hollywood blockbusters. In 2018, 24 of the top 25 highest-grossing films at the domestic box office fell under a broad category which will be called “non-original” moving forward. For the purposes of the paper, this broad designation refers to movies characterized as being franchise films (sequels, prequels, reboots, remakes, and cinematic universe entries), cinematic adaptations (based on comic books, graphic novels, literature, television shows, well-known historical events, and video games), or re-releases of older films. Among top grossers, a major shift in the sources of content—from original towards non-original ideas—has occurred since the turn of the millennium. As illustrated in the graph below, the share of non-original films among the top 25 highest-grossing films in 2000 was only 10. However, this figure has climbed in recent years, surpassing 20 in 2014 and reaching a high of 24 this past year (BoxOfficeMojo, 2019).

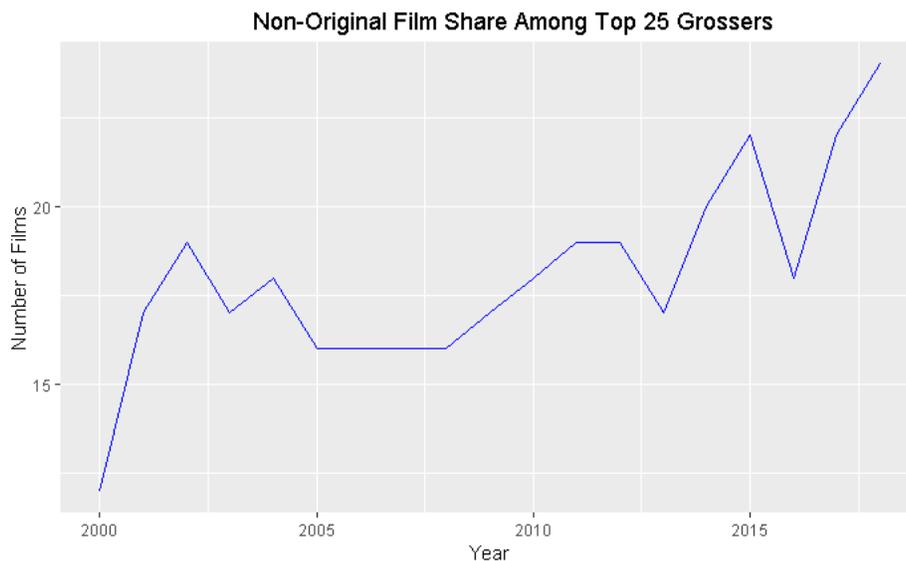


Fig 1. Share of non-original content among the top 25 highest-grossing films, 2000-2018

The trend of non-original fare comprising a larger portion of the highest-grossing domestic films suggests studios have invested increasingly more in recognizable properties in

order to maximize the revenue of their projects at the US box office. This may be due to changing consumer preferences, leading non-original films to represent the best return on investment for movie studios. The increased penetration of new technologies, such as video-on-demand, streaming services, and original content produced for online platforms, has also led to a change in the way audiences can view entertainment content and may have altered the ways studios consider assembling their portfolio of products.

Enter Netflix into the void of original popular content in American entertainment. The streaming service's arrival in the business of distributing "Netflix Originals" series and films marks a new age in the entertainment industry.¹ The first season of its first "original" series, *House of Cards*, released in February of 2013. Although still a TV show, every single episode of Season One released on the same day, allowing a viewing experience similar to television in its length, but resembling cinema in the immediate availability of each episode. Thus, the existence of the original series presents an interesting question of whether shows "streamable" in this fashion represent competition for films. Since 2013, other subscription services, such as Hulu and Amazon Prime, have similarly started to distribute their own series.

Beginning with the distribution of *Beasts of No Nation* in 2015, Netflix has also released a wide range of "Netflix Original Films" available only through streaming on the platform (Rodriguez, 2018).² This recent change should generate significant effects for the composition of new films available for consumption in theaters and through unlimited on-demand services moving forward. An even more recent development, Netflix announced in late October that it would be releasing two art-house films, *Roma* and *The Ballad of Buster Scruggs*, to limited

¹ In this case, "Netflix Originals" refers to content being distributed for the first time online, rather than on TV or in theaters.

² The same idea applies for Netflix Original films that does for Netflix Original shows.

theaters, followed shortly after by a streaming release online. The decision likely presents an effort to ensure their projects compete during awards season, rather than an attempt to maximize film revenue (Richwine, 2018). Regardless, through its original content and through its distribution of previously aired and released shows and movies, providers like Netflix are now competing with the theater industry as an alternative source of entertainment for consumers.

Thus, this paper will attempt to answer one main question: why has originality decreased in the traditional Hollywood studio system? The issue remains a pertinent question for content producers moving forward, who will continue to look for ways to generate revenue through titles which appeal to widespread audiences. In the case of Hollywood studios, perhaps non-original content will remain the only viable investment for studios looking to profit off of their films. Meanwhile, subscription-based providers, which do not need individual features to be successful, but rather need to offer a bundle of goods to entice a subscriber, may always fare better in the space for original content.

To investigate this issue, the paper will attempt to see if a change in the success of original content and non-original content has diverged as a result of the entry of new distribution models for entertainment, such as the original series or the original film, into the industry. Due to the relative recent introduction of original streaming content, the paper also hopes to motivate further investigation into the field as original online-distributed projects increase in number and spread across different sources.

The remainder of the paper will follow accordingly; Section II discusses the literature surrounding consumer purchasing theory and how this may influence audience behavior when selecting films to watch, as well as proposes two hypotheses which motivated the research; Section III discusses how the yearly non-original film and box office share among top grossing

films has changed as a result of key innovations in online distribution; Section IV discusses how the performance of individual non-original and original films has changed over time as a result of these same innovations; and Section V discusses a method to compare between the relative success of traditionally-distributed theatrical releases and online streaming-based films, as well as how non-original versus original content performs in each of these two distribution models.

II. Originality and Transitioning Distribution Models

How does the acceleration of digital distribution of entertainment relate to a shift in original content from movie studio releases to subscription-based providers? To begin with a favorite quote within the movie industry, William Goldman once said, “[N]obody knows anything” when referring to the inability of movie producers to predict whether their projects will be successful (Waldfoegel, 2017). Consequently, the trend towards a greater share of non-original content in theatrical releases may represent a risk-averse approach by studios to achieve success in the face of the uncertainty Goldman describes. But what is different about online distribution that decreases the risk of releasing original content?

A simple explanation rests in varying consumer behavior for moviegoers purchasing tickets and users of streaming services. Within consumer purchasing theory, familiarity is an extremely important variable for determining behavior. Familiarity with a product encourages a greater feeling of certainty that the consumer will be satisfied with the item and gain more utility from their purchase. Certainty likewise motivates consumers to include the product in the set of items they consider purchasing, therefore increasing the likelihood of purchase (Baker et al., 1986). Yet this effect may be entirely felt by the consumer before the purchase. In an earlier study, Berlyne (1970) proposed an opposing psychological effect, suggesting that the stimulus

created by experiencing an unfamiliar product provides greater impact on the receiver's satisfaction due to the irritation induced by repetition. Within the literature, this is referred to as the "novelty effect." However, this effect only arrives after, rather than before, the consumption of a product.

The concepts of the familiarity and novelty effect can apply to the choice to consume a film. For someone considering purchasing a theater ticket, the choice to consume allows for the viewing of only one film. Once purchased, the moviegoer must live with their decision, even if they decide they are not satisfied with the purchase of the film midway through viewing. Of course, one could exit the theater before the end credits, but given the sunk cost of traveling to the theater, this may represent an undesirable outcome. Thus, familiarity with a non-original product, for example a *Star Wars* film, may increase feelings of certainty within the consumer that they will enjoy the film. This concept is reflected in the literature. Dhar et al. (2012) demonstrated sequel films earn more than parent or standalone films by increasing theater attendance in the opening week of the project.³ The willingness of moviegoers to immediately consume sequels upon release suggests audiences prefer the familiarity of known titles.

On a Netflix-style platform, the ability to watch a range of unlimited movies after having paid the initial fee encourages a different approach to selecting films for the consumer. From a psychological standpoint, consumers may fear uncertainty less when choosing a product through a subscription provider, knowing they can switch off the program with no marginal cost (aside from their time spent watching) should it not fit their preferences. For this reason, Mark and Jay Duplass, two filmmakers who produce movies for Netflix, suggest the unlimited content model has garnered audiences for risky projects that weren't "likely to come in the traditional movie

³ Parent films refers to those which spawned one or more sequels, such as *Star Wars*. Standalone films refers to those which are not a part of a series or franchise.

rental model” (Rodriguez, 2018). With this in mind, the anticipation of the novelty effect posited by Berlyne (1970) may outweigh the benefits of familiarity discussed by Baker et al. (1986).

To highlight the salience of the subscription model, look no further than the recent popularity of MoviePass. The service, which originally allowed subscribers to purchase unlimited movie tickets for a monthly fee, replicates the risk-dispersing aspects of Netflix in a theater setting. For an industry which has seen little change in the way tickets have been purchased over time, MoviePass represents an intriguing development to decrease the significance of uncertainty on the consumer’s decision-making process.

From a supplier standpoint, the introduction of online distribution has similarly changed the way producers decide to develop projects. The digitization of content has created what Waldfogel (2017) describes as “the ‘long tail’ of niche products,” or projects which wouldn’t have previously come to the market. The theory suggests that the reduction of costs to distribute content through digital means makes producers likely to favor more obscure titles. When costs of digital distribution are lower, everyone can create more products. Thus, finding a product that falls within a specific niche, rather than resembling the growing majority of products available, offers a path towards attracting revenue. In this case, original content may find a new path to the market through subscription-based distributors.

The invention of HBO as a subscription-based service offers an illustrative example of the concept. As defined by Spence (1976), market failure occurs within the television industry when viewers’ willingness to pay for a program is higher than the program’s cost of production, but the revenue to be gained from advertising to this base is lower than the cost of production. For example, a small dedicated group of *Firefly* fans may have intensely loved the show, but the network would be unwilling to air the show because advertisers would not have wanted to sell to

such a small audience. Were the show to be picked up by a subscription-based service, such as HBO, the fans could pay the direct cost of subscription, which would be greater than the per-viewer fee paid by advertisers, with the specific intention of watching *Firefly*. The subscription-based model thus solves the market failure defined by Spence (1976).

Continuing with Netflix as an example of a subscription-based model, the service offers up an “unlimited subscription bundle,” for which customers pay monthly fees for the right to watch any of the television and movie offerings at their leisure (Hiller, 2017). Compared to the traditional theater model, in which a product must intrigue a customer into purchasing the right to watch a specific film, Netflix’s bundle must entice a customer to initially subscribe to its content and remain a subscriber (2017). Thus, the subscription model transforms behavior in the field. Since Netflix pays the licensors or producers of content a one-time fee, rather than a per-use fee (as seen on platforms like Spotify, which pay artists per listen), the popularity of one individual movie matters less than the popularity of a combination of films. Likewise, Netflix may stand to benefit more from producing three \$50 million original movies than one \$150 million superhero blockbuster, even though the same logic would fail in the theatrical release scenario.

Consequently, when examining how originality in film impacts the success of a film, this paper will attempt to prove two hypotheses. First, over time, the positive impact of non-originality on the financial success of content distributed through the traditional theater model has increased. Dhar et al. (2012) found that between 1983 to 2008, sequel films generated an increasingly higher first-week attendance (a figure which yields a strong effect on overall profitability) than did non-sequel films. Likewise, this paper hopes to build upon the work of Dhar et al. (2012) and attempt to see if significant increases in the “sequel effect,” expanded to

all non-original films, are identifiable since the recent advent of “Original” subscription-based content.

The second hypothesis posits originality in content distributed through a subscription-based service generates a larger positive (or less negative) effect on the success of the project than originality does for content distributed through the traditional theater model. As discussed, the dynamics of the subscription-based model yield greater incentives for consumers and producers alike to pursue original content. However, due to the lack of revenue attached to individual items distributed through the subscription model, identifying a metric for success poses a difficult challenge.

III. Yearly Non-Original Film Share and Box Office Share Among Top Films

3.1: Outline

The research employs four different tests related to the two hypotheses discussed in the previous section, all hoping to answer why originality in traditionally-released film has decreased over time. This section includes the first two tests to explore the ideas set forth by the first hypothesis—that non-original content has performed increasingly well at the box office in comparison to original content in recent years. Test One and Two utilize market-level data regarding non-original film share and box office share to explore this hypothesis.

For the initial test, the yearly non-original share of the highest-grossing films in the US is regressed over several time indices for dates marking innovations in online distribution. A very similar second test likewise regresses the yearly non-original share of box office revenue among the highest-grossing films over the same time indices as the first test. These two tests attempt to

draw a connection between the arrival of online distribution and an increase in non-original market share and box office share.

3.2: Data for Test One and Test Two

A wide range of data regarding movies distributed within the traditional model is available to the public. TheNumbers reports up-to-date domestic box office revenues for films based on studio reports (box office revenue does not include any aftermarket sales, such as rental or DVD sales). All of the tests for this paper focus only on the domestic North American box office. This removes factors such as staggered international release schedules and differing tastes across markets from the experiment. Within the industry, reports about film revenue are regarded as accurate.

The dataset assembled for Test One and Test Two includes the top 100 highest-grossing films at the domestic box office every year from 2000 to 2018. Utilizing the turn of the millennium as a starting point should prove sufficient for the purposes of the research. During the year 2000, the Dot-Com Boom of the mid-to-late 1990s was drawing to a close, and the availability of alternate sources of entertainment (aside from television) began to emerge thereafter. Thus, designing the experiment to encompass 2000 to 2018 should cover a time period in which new technology was always present to influence consumer choices, which will be reviewed in the next section.

Within the dataset, movies are categorized as being non-original utilizing a binary variable. To help categorize films, TheNumbers features a database which breaks down content utilizing source indices. The various indices (i.e. Books, Television, Comic Books) also includes “Original Screenplay,” or scripts for a film not based on pre-existing material. This category does, however, include sequels, since the script of some sequel films might not technically be

based on pre-existing content. Thus, films featured in the index were hardcoded within the research to display a binary variable for sequel if they were an original screenplay for a non-parent film (i.e. *The Empire Strikes Back*, not *Star Wars*) in a series. Any movie falling under any of the non-original source categories receives a 1 for non-originality.

3.3: T1 - Impact of Online Distribution on Non-original Film Share Among Top Grossers

Procedure

The first test regresses the yearly share of non-original films among the top grossing films in the domestic box office on time indices accounting for innovations in online distribution. The timeframe between 2000 and 2018 provides over a decade of data points during which innovations such as Netflix's subscription streaming service (2007) and the introduction of original series (2012, *House of Cards*) and original films (2015, *Beasts of No Nation*) emerged. Within these time frames, a significant change in the market share of non-original films over time could be expected. Consequently, the dates will remain important for investigation throughout the remainder of the paper. Test One explores whether the number of non-original films among the highest-grossing films each year between 2000 and 2018 has increased, as well as if the rate has changed following the innovations in distribution methods mentioned above.

An equation for this procedure is provided below.

Non – Original Share_t

$$= \beta_0 + \beta_1 Time_t + \beta_2 X_{2007.to.2011}_t + \beta_3 X_{2012.to.2014}_t + \beta_4 X_{2015.to.2018}_t + \epsilon_t$$

Fig 2. Regression equation for Test One.

Time_t is an indicator which starts at 0 in the year 2000, and increases by 1 every year until 2018. *X_{2007.to.2011}_t* is a binary variable which is 1 if the year in question is between

2007 and 2011. An identical rule functions for the other two time-range variables for 2012 to 2014 and 2015 to 2018.

Results

<i>Dependent variable:</i>		
Non-original Share of Top 25 Grossers		
	(1)	(2)
Time	0.354*** (0.078)	0.491*** (0.141)
X2007.to.2011		-2.135 (1.216)
X2012.to.2014		-2.632 (1.718)
X2015.to.2018		-1.930 (2.147)
Constant	14.863*** (0.825)	14.915*** (0.903)
Observations	19	19
R ²	0.547	0.645
Adjusted R ²	0.520	0.543
Residual Std. Error	1.869 (df = 17)	1.822 (df = 14)
F Statistic	20.501*** (df = 1; 17)	6.357*** (df = 4; 14)
Note:	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original Share of Top 51-75 Grossers		
	(1)	(2)
Time	0.149 (0.086)	0.107 (0.168)
X2007.to.2011		1.268 (1.452)
X2012.to.2014		-0.161 (2.050)
X2015.to.2018		1.077 (2.563)
Constant	12.868*** (0.911)	12.768*** (1.078)
Observations	19	19
R ²	0.149	0.222
Adjusted R ²	0.099	-0.0001
Residual Std. Error	2.065 (df = 17)	2.175 (df = 14)
F Statistic	2.973 (df = 1; 17)	1.000 (df = 4; 14)
Note:	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original Share of Top 26-50 Grossers		
	(1)	(2)
Time	0.191* (0.105)	0.098 (0.205)
X2007.to.2011		-0.317 (1.771)
X2012.to.2014		-0.043 (2.501)
X2015.to.2018		2.231 (3.127)
Constant	13.858*** (1.111)	14.433*** (1.315)
Observations	19	19
R ²	0.162	0.234
Adjusted R ²	0.113	0.015
Residual Std. Error	2.518 (df = 17)	2.654 (df = 14)
F Statistic	3.287* (df = 1; 17)	1.067 (df = 4; 14)
Note:	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original Share of Top 76-100 Grossers		
	(1)	(2)
Time	0.002 (0.114)	0.060 (0.228)
X2007.to.2011		0.482 (1.972)
X2012.to.2014		-0.089 (2.785)
X2015.to.2018		-1.327 (3.482)
Constant	12.405*** (1.204)	11.982*** (1.464)
Observations	19	19
R ²	0.00001	0.035
Adjusted R ²	-0.059	-0.241
Residual Std. Error	2.729 (df = 17)	2.955 (df = 14)
F Statistic	0.0002 (df = 1; 17)	0.126 (df = 4; 14)
Note:	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

T1 (top-left), T2 (top-right), T3 (bottom-left), T4 (bottom-right): Number of top grossing, non-original films falling within four ranking brackets regressed on time indices. Standard errors displayed in parentheses. Coefficients indicate number of films within rankings brackets that are non-original.

Discussion

Tables 1-4 analyze this question in two distinct ways, with Table 1, 2, 3, and 4 isolating the effects of time on the non-original share among the top 25, top 26 to 50, top 51 to 75, and top 76 to 100 highest-grossing films, respectively. In the first regression (left column) in all tables, the yearly number of non-original films among the top grossers is simply regressed on time, an indicator variable which climbs by one each year. In the second regression in all tables, the yearly non-original share is regressed on time and indicators for 3 time periods, indicating whether the film released between 2007 to 2011, 2012 to 2014, and 2015 to 2018.

When analyzing the first regression, the yearly time variable generates a substantial impact on non-original prevalence among top-grossers in Table 1. We can interpret this result as saying the amount of non-original films is expected to rise by 0.354 films per year between 2000 and 2018 beginning with 14.9 films in 2000. While incremental each year, this represents a total increase of over 6 films by 2018, a sizable amount for a set containing just 25 films. The expectation of over 21 films being non-original in 2018 among the top 25 grossers approaches but does not quite reach the actual amount of 24 observed.

Similarly, the amount of non-original films within the top 26 to 50 grossers increases at both a sizeable and statistically significant rate of 0.191 films per year, representing a total expected increase of about 3.5 films by 2018. In Tables 3 and 4, however, the first regression yields statistically insignificant results, as well as increasingly smaller effects which fall to zero by the last table. Therefore, while offering nothing conclusive, especially given the small sample size of this test, the first regression in Tables 1 through 4 illustrates the general increase in non-originality over time among the highest rankings brackets (T1 and 2), and the absence of this trend among the lower ranking brackets (T3 and 4).

Moving on to the second regression, the results proved drastically less statistically significant. This is attributable to the low number of observations within each time period bracket, which can be as low as 3. When interpreting these results, the number of non-original films remains greatest in the higher of the four rankings brackets within almost all of the respective time periods. Compared to the first regression, these results offer a slightly different approach by showing the rate of increase is greater in later years, especially within the first two tables. For example, utilizing the results for the second regression in Table 1, we would expect the non-original share to be 14.92 films in 2000, 16.22 films in 2007, 18.18 in 2012, and 20.35 in 2015. Although a clear tendency towards large increases occurring in the latest time period appears in Table 1 and Table 2, a lack of any recognizable pattern begins to emerge within the results for Table 3 and Table 4. It is again imperative to stress the lack of statistical significance present in these results, likely a result of the small sample size (19 yearly data points).

Test One illustrates some of the changes occurring within the film industry. The number of non-original films is increasing, especially among the highest of the top-grossing brackets, and especially within recent years. Intuitively, this makes sense; non-original films attract the largest audiences, and thus occupy higher positions within the box office. Test Two follows up the findings of Test One by asking how the financial incentives of releasing non-original content is influencing the proliferation of non-original content among top-grossers at the US Box Office.

3.4: T2 - Impact of Online Distribution on Non-original Box Office Share Among Top Grossers

Procedure

An alternative approach which may offer more relevant findings focuses on the financial success of non-original films, which incorporates both industry incentives and the magnitude of financial success of the films at different rankings within the box office. Thus, while Test 2

follows a similar methodology to Test 1, the dependent variable is changed to the percentage of the total box office of top grossers captured by the revenue of non-original films. For example, if five of the ten highest-grossing films in 2000 were non-original films, and these five earned 5 million dollars, and in total the ten highest-grossing films earned 10 million dollars, the non-original share of the box office in 2000 would be 50 percent. Like before, for the first and second regressions in Tables 5-8, non-original revenue share is regressed on the yearly time index variable and the binary variables for 2007-2011, 2012-2014, and 2015-2018.

An equation for this procedure is provided below.

$$\begin{aligned} & \mathbf{Non - Original\ BO\ Share}_t \\ &= \beta_0 + \beta_1 \mathbf{Time}_t + \beta_2 \mathbf{X2007.to.2011}_t + \beta_3 \mathbf{X2012.to.2014}_t \\ &+ \beta_4 \mathbf{X2015.to.2018}_t + \epsilon_t \end{aligned}$$

Fig 3. Regression equation for Test Two.

The only variable different in the regression equation for Test Two is the dependent variable, non-original box office share, which is expressed as a percentage rather than in absolute numbers. Revenue for films has been adjusted to be in 2018 US dollars to prevent biases caused by inflation within the data.

Results

<i>Dependent variable:</i>		
Non-original BO Share of Top 1-25 Grossers		
	(1)	(2)
Time	0.014*** (0.003)	0.019*** (0.006)
X2007.to.2011		-0.071 (0.055)
X2012.to.2014		-0.088 (0.077)
X2015.to.2018		-0.075 (0.096)
Constant	0.628*** (0.035)	0.628*** (0.041)
Observations	19	19
R ²	0.511	0.573
Adjusted R ²	0.483	0.451
Residual Std. Error	0.079 (df = 17)	0.082 (df = 14)
F Statistic	17.793*** (df = 1; 17)	4.701** (df = 4; 14)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original BO Share of Top 26-50 Grossers		
	(1)	(2)
Time	0.008* (0.004)	0.002 (0.008)
X2007.to.2011		-0.011 (0.070)
X2012.to.2014		0.022 (0.098)
X2015.to.2018		0.116 (0.123)
Constant	0.554*** (0.044)	0.583*** (0.052)
Observations	19	19
R ²	0.169	0.262
Adjusted R ²	0.120	0.052
Residual Std. Error	0.100 (df = 17)	0.104 (df = 14)
F Statistic	3.460* (df = 1; 17)	1.246 (df = 4; 14)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original BO Share of Top 51-75 Grossers		
	(1)	(2)
Time	0.007* (0.004)	0.005 (0.007)
X2007.to.2011		0.060 (0.061)
X2012.to.2014		-0.004 (0.087)
X2015.to.2018		0.048 (0.109)
Constant	0.504*** (0.039)	0.498*** (0.046)
Observations	19	19
R ²	0.176	0.258
Adjusted R ²	0.128	0.046
Residual Std. Error	0.088 (df = 17)	0.092 (df = 14)
F Statistic	3.635* (df = 1; 17)	1.215 (df = 4; 14)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

<i>Dependent variable:</i>		
Non-original BO Share of Top 76-100 Grossers		
	(1)	(2)
Time	-0.0002 (0.004)	0.003 (0.009)
X2007.to.2011		0.010 (0.077)
X2012.to.2014		-0.001 (0.108)
X2015.to.2018		-0.068 (0.136)
Constant	0.502*** (0.047)	0.484*** (0.057)
Observations	19	19
R ²	0.0001	0.042
Adjusted R ²	-0.059	-0.231
Residual Std. Error	0.107 (df = 17)	0.115 (df = 14)
F Statistic	0.002 (df = 1; 17)	0.155 (df = 4; 14)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	
Time	2007-2011	2012-2014 2015-Present

T5 (top-left), T6 (top-right), T7 (bottom-left), T8 (bottom-right). Share of total domestic box office by non-original content falling within four ranking brackets regressed on time indices. Standard errors displayed in parentheses. Coefficients indicate non-original share of box office as a percentage within rankings brackets.

Discussion

Within the first regression, the coefficients for the time index in Tables 5 through 7 offer a significant result. Regression 1 in Table 5 shows to expect the non-original box office share to increase by 1.4 percentage points each year from 2001 to 2018 among the top 25 films, beginning with about 63 percent in 2000. Likewise, Tables 6 and 7 show to expect an increase of about 0.8 percentage points each year for the next two rankings brackets beginning with 55 percent and 50 percent, respectively. This effect has disappeared by Table 8. To contextualize the rate of that growth, we'd expect the non-original box office share among the top 25 grossers to be over 25 percentage points higher last year than it was at the turn of the millennium, and about 14 percentage points higher for the next two rankings brackets over the same period. This growing trend in the financial dominance of non-original content at the domestic box office supports the paper's hypothesis that a relationship exists between financial success and non-originality among top-grossing movies which is increasing over time.

But does this relationship change due to the key innovations introduced over time? Again, the second regression does not yield statistically significant results. It's important to note many of these results are not robust given the small sample size. Like when the regression equations were carried through in Test One, the results of the second regression show a familiar trend; the financial impact of non-originality on the domestic box office is growing over time, and carries the highest influence in the higher rankings brackets before disappearing in the lowest bracket. Adjusted for the length of the time periods, the rate of growth of non-original box office share is fastest during the last time period among the 2 highest rankings brackets, showing that the financial success of non-original content is increasing in recent years.

Given the lack of statistical significance in the second regression, not much—if anything—can be concluded from Test One or Test Two. What the tests do exhibit is the importance of the last time period, 2015 to 2018, in proving the validity of Hypothesis One. The yearly statistics utilized in the first two experiments may offer a high-level picture of the trends occurring in the movie industry. But with only 19 dependent variables to choose from—the yearly number of top grossers and the yearly box office share—the analysis is subject to variations which can exist in such a small sample size. To expand the scope of the investigation, Test Three analyzes the individual financial success of each film based on its characteristics, including non-originality, in order to confirm a shift is occurring during this final time period.

IV. Individual Film Revenue as a Function of Non-originality

4.1: Outline

This section includes Test Three, which will further explore the ideas set forth in the Test Two by investigating the financial success of the individual films among the yearly top 100 highest-grossing films, rather than focusing on trends from an industry-level perspective. For this test, the real revenue of a film is regressed on a number of characteristics, including non-originality, release month, critical rating, genre, and film profile, to analyze how the source material of a film impacts the financial success of the project. The time indices will also be included to assess how these effects have changed over time.

4.2: Data for Test Three

Test Three utilizes an expanded version of the dataset for Test One and Test Two. Like the first dataset, this dataset features the top 100 highest-grossing films for each year between 2000 and 2018 categorized by their source type and having an overall binary variable for non-

originality. Again, this dataset also includes indicator variables for time and binary variables for whether a film falls within a specific time bracket (2007 to 2011, 2012 to 2014, and 2015 to 2018) to represent innovations in online distribution. However, Test Three analyzes the films on an individual level, so information regarding the film's profile, genre, critical rating, release month, and MPAA rating are included in the data to provide appropriate control variables.

A movie's budget offers an indication of the size of a project's "profile," or how well-known it is among consumers prior to release. Thus, budget data could serve as an important control when comparing between projects with disparate profiles, such as *Star Wars* and *Creed*, which are both non-original projects. In the case of *Star Wars*, Disney is willing to invest a budget in the hundreds of millions because of the large scale fanbase for the property. Although still popular, MGM knows the *Creed* franchise has less fans than its intergalactic counterpart, and thus won't invest as much towards the film's budget. While TheNumbers features a database of budgets, it also includes a warning that budget data is often kept secret or manipulated by studio executives. Additionally, the number of budgets available within these indices proves a limiting factor for using budget as a control variable. Therefore, budget was not used in the test.

To compensate for the lack of a control variable for film profile, the number of critical reviews can be used as an alternative indicator, similar to how Duan et al. (2008) employed user ratings as a proxy for sales. Film critics work for publishers who want to feature movie reviews which their audience will be interested to read. With this logic, more critical reviews of a larger scale film, like *Star Wars*, will appear in the buildup to the film's release compared to a mid-tier project like *Creed*. For reference, the two most recent films in both franchises, *The Last Jedi* and *Creed II*, have 414 and 208 critic reviews posted on Rotten Tomatoes, respectively. Thus, the quantity of critical reviews for a project on Rotten Tomatoes will serve as a control variable for a

film's profile prior to its release. Profile as a control variable is also important for distinguishing original content which may have an external factor lending it more visibility to the public. The presence of a famous director at the helm, such as James Cameron, makes a film a huge property, despite its lack of attachment to any pre-existing material.

TheNumbers includes several other indices which helped provide the control variables necessary to isolate the originality factor within the research. TheNumbers includes genre indices, which incorporates 11 different film genres. Like film profile, genre can be used to control for the fact that audiences are interested in different types of projects. An action movie such as *The Fast and the Furious* appeals to a larger movie-going crowd than does a dramatic movie like *Moonlight*. Other control variables, such as release date, MPAA rating, and critical ratings, are readily available within the indices on TheNumbers and Rotten Tomatoes. With the availability of ratings through the internet, audiences can see how much critics liked a movie, which is an important influence on their decisions to go see a film. Likewise, Rotten Tomatoes critical score (percent approval) is included as a control variable in the dataset. A binary variable for each month of the year provides an important control given films experience more success in various seasons, such as during summer or the holidays. Last, the MPAA rating (i.e. G, PG, PG-13, R, etc.) is included as a control to account for the different size and ages of audiences who may attend a film based on its MPAA rating.

4.3: T3 - Box Office Revenue of an Individual Film Based on Non-originality

Procedure

Unlike the first two tests, Test Three regresses the revenue of an individual film (originating from the dataset of the top 100 grossing films each year between 2000 and 2018) on the non-originality variable to assess how source material affects financial success. Thus, non-

originality as a binary independent variable (1 for non-original) is added to the experiment, and the dependent variable is the real domestic box office revenue of the film (in millions of 2018 US dollars). Additionally, the real revenue is regressed on interaction terms which indicate whether a non-original film released during one of the specified time periods—2007 to 2011, 2012 to 2014, and 2015 to 2018—representing the key innovations in online distribution (i.e. Netflix’s online streaming, the release of Netflix Original shows, and the release of Netflix Original Movies). 1900 films are included in the analysis.

An equation for this procedure is provided.

Real Revenue_i

$$\begin{aligned}
 &= \beta_0 + \beta_1 \text{Non-Original}_i + \beta_2 \text{Time}_i + \beta_3 \text{RTScore}_i + \beta_4 \text{RTNumRevs}_i \\
 &+ \sum_j \beta_j \text{Time Brackets}_i + \sum_k \beta_k \text{Non-Original}_i * \text{Time Brackets}_i \\
 &+ \sum_h \beta_h \text{MPAA Rating}_i + \sum_l \beta_l \text{Genre}_i + \sum_g \beta_g \text{Month}_i + \varepsilon_i
 \end{aligned}$$

Fig 4. Regression equation for Test Three.

Real Revenue is expressed in millions of US dollars. Non-Original_i, Time_i, and the controls for the three time period brackets follow the same rules as in Section III. An interaction term for the time brackets is also included (i.e. a non-original film from 2016 will receive a 1 for Non-Original_i, X2015.2018_i, and Un2015.2018_i). RTScore_i and RTNumRevs_i are continuous control variables representing a film’s Rotten Tomatoes critical score (percentage of critics who enjoyed the film) and the number of critical ratings. There are 6 MPAA ratings, 11 genres, and 12 release month binary variables that are 1 if the film falls within those categories.

Results

<i>Dependent variable:</i>				
	Real Rev (MM USD)			
	(1)	(2)	(3)	(4)
Unor	35.065*** (4.597)	35.215*** (4.613)	24.491*** (7.487)	10.025 (6.140)
Time		-0.171 (0.412)		
X2007.2011			1.230 (8.966)	-7.376 (7.303)
X2012.2014			-7.318 (10.686)	-20.180** (8.680)
X2015.2018			-25.498** (10.011)	-51.730*** (8.196)
Un2007.2011			1.637 (11.667)	-5.829 (9.410)
Un2012.2014			16.312 (13.801)	1.523 (11.171)
Un2015.2018			37.927*** (12.694)	10.890 (10.294)
PG				30.352*** (6.360)
PG.13				33.040*** (4.216)
R				-15.985** (6.943)
Not.Rated				88.978* (46.139)
May				39.154*** (7.303)
Jun				24.922*** (6.491)
Sep				-20.080*** (7.223)
Dec				14.082** (5.754)
Action				34.252*** (5.418)
Adventure				61.780*** (5.653)
Concert.Performance				50.699 (32.586)
Horror				20.262*** (7.350)
RT.Score				0.254*** (0.087)
RT.NumRevs				0.619*** (0.036)
Constant	86.780*** (3.569)	88.233*** (4.994)	92.352*** (5.674)	-48.477*** (7.157)
Observations	1,900	1,900	1,900	1,900
R ²	0.030	0.030	0.035	0.384
Adjusted R ²	0.029	0.029	0.032	0.377
Residual Std. Error	98.066 (df = 1898)	98.088 (df = 1897)	97.943 (df = 1892)	78.568 (df = 1878)
F Statistic	58.172*** (df = 1; 1898)	29.160*** (df = 2; 1897)	9.872*** (df = 7; 1892)	55.696*** (df = 21; 1878)
Note:				*p<0.1; **p<0.05; ***p<0.01

T9: Real revenue regressed on non-originality and time indicators. Coefficients signify millions of USD. X variables (i.e. X2007.2011) indicate original films released between the specified dates, while Un (Un2007.2011) variables are interaction terms for non-original films released between the specified dates. The last regression includes all relevant control variables (many were not presented in the table due to lack of significance).

Discussion

Due to its greater sample size, Test Three yielded the most robust results thus far. Without any time-related controls, the first regression indicates we should expect an original film to gross about \$86.8 million at the box office, while we would expect that figure to climb by \$35.1 million if the film were non-original. On an economic level, this is an extremely significant result. Simply distributing a movie that is non-original will increase the expected profits of the product by over 40 percent within this regression. The coefficient for non-originality is significant at the one percent level, demonstrating a strong relationship between a film's source material and its financial success. The second regression, which incorporates a time index similar to that in the first two experiments, shows the time coefficient is not significant, even at the 10 percent level. Thus, non-originality, and not time is explaining the difference in revenues. This result makes logical sense given revenues are inflation-adjusted.

The third regression, however, offers more insight into the recent financial advantage non-original films have enjoyed at the box office. This equation shows we should expect original films released between 2000 and 2006 to gross \$92.4 million, with that number rising by \$24.5 million if the film were non-original (significant at the 1 percent level) during the time period. An additional statistically significant change in the film's financial success based on source material occurs for films released between 2015 and 2018; original films stand to make \$25.5 million dollars less than they would have between 2000 and 2006, while non-original films stand to make \$38.9 million dollars more. In sum, a non-original film released between 2015 and 2018 stands to earn about \$62.4 million dollars more than an original film released during the same time period, an economically significant increase of 93 percent. Even when compared to a non-original film from 2000, the expected increase in real revenue for a non-original film released between 2015 and 2018 stands at 10 percent.

While the creation of films released solely through online streaming platforms in 2015 marks an important innovation in the film industry, it does not signify that this specific date is the cause of this shift. However, it does appear that the divergence truly emerges around this point in time. Similar to Test One and Two, in which the 2015 time period offered the most sizable results, a notable difference arises in box office trends surrounding this date, particularly for non-original content.

When controlling for other variables which might have an impact on a film's financial success in the fourth regression, the marginal effect of non-originality in the 2000 to 2006 time period shrinks to \$10 million—still an increase of almost 20 percent—with a significance just outside the 10 percent level. The coefficients for time period, however, remain high in the final two time periods. This signifies an original film released between 2015 and 2018 is expected to earn \$51.6 million less than an original film released between 2000 and 2006 (controlling for the other variables), an effect significant at the 0.001 percent level and a notably sizable impact. The financial success of individual original films is expected to decrease by over 100 percent during the final time period, demonstrating a rapid transition in which originality harms the financial success of films much more in recent years.

Interestingly, the interaction term for non-originality during this time period does not offer significant results. This might suggest that while original films suffer, the gains experienced by non-original films are captured instead by the other control variables. Among non-original films, a prevalence for franchise films exists, which could lead expected economic performance to be represented by the control variables often associated with these types of films. These may include PG-13 MPAA ratings, early summer or holiday release dates, and the action or adventure genre. This idea will be explored in the subsequent sub-section.

It is important to note certain control variables, including the ones mentioned above, yield an enormous economic effect on the financial success of a film, oftentimes to a greater extent than the variables accounting for the time period. Adventure films, for example, are expected to yield \$61.8 million more in revenue than the baseline film, while films released in May are expected to yield \$39.2 more in revenue. Both of these factors are significant at the 0.001 percent level.

Rotten Tomatoes scores also pose a subtle financial effect for films of disparate critical quality. For example, a film receiving a 0 percent score on the website is expected to earn \$25.4 million less than a film receiving a 100 percent score. For the film with the median Rotten Tomatoes Score, 53 percent, the regression shows to expect the revenue to be \$5.8 million dollars more than the film with the score at the 25 percent quartile (30 percent), and \$5.8 million dollars less than the film with the score at the 75 percent quartile (76 percent). Since Rotten Tomatoes can only vary by 100, the relative effect of Rotten Tomatoes scores remains small when compared to other controls, despite its statistical significance.

Film profile, however, can pose a much greater economic effect. For every 100 more reviews a film received on Rotten Tomatoes, the revenue is expected to climb \$61.9 million. Since the amount of reviews within the dataset can vary by several hundred in number, this yields perhaps the most significant increase in financial success. To illustrate, *Avatar*, which received 303 reviews, is expected to earn \$175.2 million more than *Fireproof*, which received 20 reviews (holding all other controls constant). *Avatar*, possessing a greater name recognition due to having James Cameron as director, obviously received a lot of attention, giving it a higher profile and helping it garner more financial success than most original films. Films falling within the middle two quartiles of film profile should expect less, but still sizable, deviation due to this

variable. The 25th and 75th percentile are only separated by 92 reviews (123 to 215), meaning film profile accounts for a \$56.9 million increase between these two quartiles.

Critical review count remains an imperfect measure of how to assess the profile of a film due to its crossing-over in many instances with the originality category. Additionally, given the proliferation of multimedia news outlets, modern films have more reviews available on Rotten Tomatoes than older films in the dataset. *Avatar*, regardless of being the highest grossing international film of all time, has over 100 less reviews than *Black Panther*. Likewise, when holding all other controls constant, profile is not the most effective variable to analyze in isolation.

Revenue by Source Content

<i>Dependent variable:</i>			
Real Rev (MM USD)			
		PG	3.165 (10.813)
Book.	4.787 (5.160)	PG.13	-4.097 (11.162)
Comic.Book.	29.271*** (8.892)	R	-39.921*** (11.498)
TV.	8.263 (8.602)	Not.Rated	47.024 (45.042)
Remake.	4.586 (7.034)	May	30.959*** (7.028)
Play.	2.762 (13.793)	Jun	19.964*** (6.240)
Spin.off	27.721* (15.204)	Sep	-13.440* (6.932)
Video.Game.	-0.434 (18.894)	Dec	19.032*** (5.556)
History.	-16.194** (8.240)	Action	16.725*** (5.575)
Re.release.	-33.486 (23.034)	Adventure	40.187*** (5.953)
Sequel.	60.493*** (5.288)	Concert.Performance	36.303 (31.272)
X2007.2011	-2.825 (6.647)	Horror	7.359 (7.249)
X2012.2014	-14.229* (8.019)	RT.Score	0.341*** (0.086)
X2015.2018	-44.647*** (7.547)	RT.NumRevs	0.624*** (0.036)
Un2007.2011	-18.781** (8.164)	Constant	-18.389 (11.855)
Un2012.2014	-14.297 (10.216)	Observations	1,900
Un2015.2018	-10.931 (9.264)	R ²	0.440
		Adjusted R ²	0.431
		Residual Std. Error	75.066 (df= 1869)
		F Statistic	48.985*** (df= 30; 1869)
		<i>Note:</i>	* p<0.1; ** p<0.05; *** p<0.01

T10: Real Revenue in millions of 2018 USD regressed on each category for non-originality, yearly indicators, interaction terms, and relevant control variables. Coefficients are in millions of 2018 USD. Table is one single column but presented side by side due to its length.

In the fourth regression of Table 9, the coefficients for non-originality and the interaction terms for time period and non-originality lost their statistical significance. As briefly discussed, other control variables aside from or associated with non-originality, like MPAA rating, genre, or release month, may have influenced the lack of significance. To explore this idea, the films included in the dataset were also categorized by their specific source content, rather than just an

overall binary variable for non-originality. These sources included books, comic books, TV shows, remakes, plays, spin-offs, video games, historical events, re-releases, and sequels.

When real revenue is regressed on these categorical variables, unsurprisingly, films derived from comic books or sequel films are expected to earn the most at the box office. As mentioned before, many of these films are high-budget superhero or franchise films, which rank among some of the most successful titles today. It is therefore unsurprising that Table 10 shows comic book movies should be expected to gross \$29.3 million more than the baseline original film, while sequel films should be expected to gross \$60.5 million more. Both are statistically significant results. Meanwhile, the coefficients for MPAA rating, release month, and genre all shrink in magnitude compared to those in Table 9, showing these control variables were earlier capturing some of the effect of these specific categories of non-originality. This phenomenon offers a probable explanation for the drop in size and statistical significance for the final interaction term in the fourth regression of T9, given much of the expected earnings of a project is tied up in the type of non-originality, rather than just non-originality itself.

When accounting for these controls in the fourth regression, the effect of originality on the individual success of the top 100 highest-grossing films at the domestic box office remains pertinent. Likewise, Test Three supports the ideas of the paper; non-original films released through the traditional distribution model have seen a rise in financial success during the most recent time period, even when accounting for other factors which determine how well a film will sell in theaters.

Combined with the results of the first two tests, which isolated the years between 2015 and 2018 as a potential time period in which this shift occurred, it is entirely plausible that the innovations in online distribution of entertainment content have altered consumer preferences

and generated the increasing success of non-original films. This could have occurred due to the gradual integration of innovations like online streaming (introduced by Netflix in 2007) or Original shows (2012) into the mainstream, the recent entrance of Original streaming films (2015), a combination of all three innovations, or a different external factor. Thus, the success of non-original versus original content in the online distribution model will be explored further in Section V to further investigate why originality has decreased in the traditional film industry.

V. Test Within the Non-Traditional Distribution Model

5.1: Outline

While Tests One, Two, and Three all investigated the first hypothesis, Test Four explores the second hypothesis: non-originality of subscription-based streaming content yields a smaller effect on the success of the project than does non-originality for content distributed through the traditional theater model. This section of the paper moves beyond current research and hopes to begin a conversation about how the types of content which will appear in the future through the different distribution models will continue to evolve. Test Four incorporates a different dataset utilizing electronic word of mouth in place of revenue to compare between streaming-based and traditional content.

5.2: Theory

One of the main issues in examining the success of original content in the subscription-based setting is the lack of a revenue stream directly related to the product. When a consumer pays for HBO, they may only be watching *Game of Thrones*. Yet their monthly fee doesn't directly represent revenue for the one program, but rather the whole platform. Even more difficult to analyze is the lack of viewing statistics for the streaming-based subscription services,

like Netflix and Hulu (HBO does release “ratings” similar to a traditional TV channel when a film debuts on the premium channel). Although third party estimates of “ratings” for original shows and movies, such as *Stranger Things* and *Bright*, do exist, Netflix’s own Chief Content Officer has publicly stated the estimates are inaccurate (Koblin, 2017). For streaming services, viewing numbers represent extremely valuable proprietary data, offering insight into the type of programming which customers will enjoy, and are thus unavailable in a precise form.

Analyzing electronic word of mouth (eWOM) as a substitute for actual revenue or viewing statistics may offer an alternative path to assess the relative success of subscription-based content. In the literature regarding word of mouth (WOM), or interpersonal communication, researchers have always considered WOM an important factor for consumers when choosing whether to purchase a product. Katz and Lazarsfeld (1955) showed WOM was the single most important source of influence for consumers choosing between household items, and Slack (1999) asserted people choose to visit a new website because of a personal recommendation 57 percent of the time. However, studies of the impact of WOM on sales stalled due to the difficulties associated with recording all verbal conversation. In response, Godes and Mayzlin (2004) proposed the study of eWOM, or online discussion of a product. On the internet, communications are more easily recorded and traceable, solving the problem associated with a lack of records for conversations.

Interestingly enough, two influential papers for eWOM, Godes and Mayzlin (2004) and Duan et al. (2008) both utilized the movie industry as a topic for their research, since the industry had always received significant academic attention for studying the effects of WOM. Godes and Mayzlin (2004) concluded their data suggested the greater the volume of eWOM for a film, the higher impact of the buzz on future sales. However, the nature of eWOM as an endogenous

variable in generating sales remains important to note in this setting. Duan et al. (2008) discuss the “positive feedback mechanism” of eWOM, noting that eWOM “is not only a driving force in consumer purchase, but also an outcome of retail sales.” Thus, when considering eWOM as a measure of success for a product, it is important to consider it as an endogenous variable while performing regression analysis. Godes and Mayzlin (2004) acknowledge the time-dependent nature of eWOM, stating although overall eWOM may generate higher aggregate sales, “high [e]WOM today does not necessarily mean higher sales tomorrow. It may just mean that the firm had high sales yesterday.”

To reflect this element, Duan et al. (2008) measured eWOM on a daily basis, stressing the importance of the short interval. With relation to the movie industry, the study first regressed the volume of eWOM on Box Office revenue, and subsequently revenue on eWOM. Through their research, they found that daily postings of user reviews displayed a strong positive correlation with the daily gross of a film. However, they also discovered that a lag in eWOM, such as user reviews posted on day $t-1$ or day $t-2$, demonstrated a weaker correlation with the daily gross of the film on day t (2008). Thus, by including lagging terms, the study was better able to pinpoint causality between eWOM one day and sales that same day. When developing a relationship regarding eWOM, employing measures from different time periods can be important for removing endogeneity.

In the years since the discussion of eWOM began, the proliferation of social media also provides a new vehicle for eWOM. While the dispersion of user reviews represented the only eWOM factor in the experiments discussed earlier, the sharing of reactions to a film via social media presents an alternative variable to analyze. Despite hesitations about the uncertainty of models utilized to predict the impact of social media buzz on box office revenue, Lehrer and Xie

(2016) propose that social media can be utilized to improve box office predictions. While their approach utilized eWOM valence, or the overall sentiment of the post, the ability to collectively measure the feelings of social media posts about a product has posed problems in experiments (Godes & Mayzlin, 2008). Analyzing the overall sentiment of eWOM towards a single title can prove time-consuming and inaccurate, while also limiting the ability of researchers to develop large datasets. Likewise, the research for this paper will instead focus on eWOM volume, and not valence, as a measure of film success.

While the concerns of Godes and Mayzlin (2008) towards eWOM valence rest in the difficulties of estimating the effects of positive versus negative eWOM, this paper will use search hits as a source of eWOM volume neutral in its valence. While a Google search, for example, does not constitute interpersonal communication, it does demonstrate an awareness of a consumer towards the existence of a product. To demonstrate, Wu and Brynjolffson (2015) developed a model by which search frequencies of housing prices in various markets were utilized to predict future housing prices. Furthermore, the authors claimed their model beat the National Association of Realtor's predictions for the period by 23 percent (2015). Although not applicable in every industry, in the case of Netflix, Piper Jaffray analysts have posited Google Trends as an indicator of general trends in subscription growth, citing a 4 percent error in the use of tracking Google hits of Netflix for predicting growth of subscribers (Wolff-Mann, 2018).

To summarize the importance of WOM, and more specifically, eWOM, to the research, the prevalence of data related to buzz about a product will serve as a substitute for revenue and viewership data of subscription-based streaming services. An analysis of eWOM utilizing more traditional sources of interpersonal communication, such as volume of critical or user reviews of films, or newer sources, such as volume of mentions over social media, or proxies for eWOM,

such as Google search volume, can provide a method for comparing the relative success of content across the two different distribution models.

5.3: Data for Test Four

The dataset features both traditionally-distributed and streaming-based content. The eWOM for each of the yearly 100 highest-grossing films at the domestic box office released after June of 2014 (over a full calendar year before the release of *Beasts of No Nation*, the first Netflix Original film) until the end of 2018 is incorporated into the new dataset, as well as the eWOM for every Original film released by Netflix from October of 2015 (release date of *Beasts of No Nation*) to April of 2019. eWOM is collected using the Google Trends web app. The Google Trends poses issues, since it only records eWOM of different search terms relative to an index, rather than giving an absolute number of search hits for a given time period. Additionally, the app only allows for five entries at a time. Thus, the approximately 450 traditional movies and 90 Netflix Original films released during the specified time periods were input five at a time before being re-indexed to the last film in the previous set.⁴

Unfortunately, due to the lengthy process involved with collecting and processing data in this manner, more films from earlier time periods were not able to be added to the set. Likewise, the research does not contain an intertemporal comparison to see how the effects of non-originality have changed over time as part of Test 4. Additionally, many films had to be removed from the set because of issues singling out the eWOM of film titles which may be common

⁴ For example, if five films were entered into the Google Trends app, the next set would feature the fifth film as the new first entry. In this manner, the trends can re-indexed every four films to allow for comparisons between each set. A Google Trends API package is available in R which could allow for more efficient bulk-entry and processing of film titles than the by-hand method utilized in the paper. The research presented here decided not to use this package due to time constraints of writing a suitable program. Additionally, the program is often blocked by Google if too many users on one Wi-Fi network run too many terms through the API. For future students, this may pose issues. Alternative collection methods are recommended.

search terms.⁵ Many one-word film titles, such as *Lucy*, generated noise unassociated with the release of the film due to other searches which may have been for alternative queries (i.e. the show *I Love Lucy*, the character Lucy from *Peanuts*, or the actress Lucy Liu). While this does introduce a possible error into the data, there is no clear directional bias due to the random nature of film titles relating to common phrases. Potential biases could stem from franchise films holding longer, more specific titles which would rarely be eliminated, favoring larger films in the dataset. After removing problematic search terms, the data-set features 341 traditional and 81 streaming-based titles.⁶

Additional errors arise due to the time periods available in the Google Trends app. Given the length of time analyzed in the paper, the indices provided in the results for each set only measured how frequently searched a term was over the course of a week. Thus, the data contains a weekly index for the week beginning on the Sunday before the release of the film, as well as a second weekly index for the week beginning on the Sunday after the release of the film. This does not allow for the same precise daily analysis utilized by Duan et al. (2008). For films which may have released on different days, this may introduce an error given the differing amounts of time before and after a film was released within the indices for the first and second week. For traditionally-distributed films, wide releases usually occur on Thursday nights, with some irregularity. For Netflix projects, however, the release day is more erratic. While this does

⁵ The app does sometimes allow to search for a movie instead of a term, but this option is not available for every title, thus leading to the dropping of troublesome titles for which this option was unavailable.

⁶ *Bird Box*, released by Netflix in December of 2018, was also removed from the dataset due to its large outlier status for the Week 2 index. Its eWOM index for Week 2 was eight times greater than the next closest Netflix title, which significantly skewed the non-originality term in Table 13. One could argue the popularity of the film had more to do with its viral status, including a popular internet meme called the “Bird Box Challenge,” than the popularity of its source material, the original novel. This type of popularity does not fit with the research and was removed accordingly.

introduce noise into the data, there is again no obvious directional bias due to the consistent release dates of theatrical films and the random release dates of Netflix films.

For future research utilizing eWOM as a means of comparison between the traditional and streaming-based distribution models, the use of a more precise data-collecting tool is recommended. The incorporation of daily lagging terms will allow for greater specificity and better ensure the elimination of endogeneity. Additionally, as Netflix and other streaming services expand their portfolios of online films, the sample size of an accompanying dataset can be expanded to allow for a more robust future investigation.

Like Test Three, the films are categorized by source material, with a collective non-original binary variable and indicators for each type of source. The control variables from Test Three (critical rating, film profile, genre, and release month) aside from MPAA rating (not available for every Netflix title) are carried over as well.

5.4: T4 - Impact of Non-originality on eWOM for Both Distribution Models

Procedure

The fourth test regressed eWOM for films on an interaction term for distribution model and originality, while controlling for a range of variables carried over from Test Three. The analysis utilized the collective volume of google searches for the weeks beginning on the Sundays before and after the release of the individual film, which will be called Week 1 and Week 2.

While imperfect, this does capture the buzz around a film during the first and second week of its release. Analyzing only the start of a film's run can be a more efficient method than incorporating the entirety of the film's stay in the box office. The importance of the opening weekend time frame is stressed by Dhar et al. (2012), due to the fact that non-original content

often shifts audience demand towards the opening-weekend of a film. While this does not ensure better “legs,” or audience retention within the industry, front-loaded films enjoy the highest profitability by amassing most of their revenue immediately. Thus, this research included an opening week and a second week and will be run with both the Week 1 and Week 2 index as dependent variables.

An equation for the procedure is provided below.

Weekly Index (1 or 2)_i

$$= \beta_0 + \beta_1 \text{Non-Original}_i + \beta_2 \text{Theater}_i + \beta_3 \text{Non-Original}_i * \text{Theater}_i + \beta_5 \text{Time}_i + \beta_6 \text{RTScore}_i + \beta_7 \text{RTNumRevs}_i + \sum_h \beta_h \text{Genre}_i + \sum_l \beta_l \text{Month}_i + \epsilon_i$$

Fig 5. Regression equation for Test Four.

Similar to before, Non-Original_i is a binary variable which will be 1 for non-original content and 0 for original content. Likewise, Theater_i is a binary variable which will be 1 for content distributed through theaters and 0 for content distributed through online streaming. This remains an extremely important control, since it will eliminate the effect of different audience size and behavior for films released in theaters versus on a Netflix or Hulu without other control variables. For example, despite the popularity of both properties, *Star Wars* may appeal to a broader audience than *Stranger Things* due to the visibility of theater releases. The third beta will be associated with an interaction term for Non-Original_i and Theater_i .

Thus, this third beta will only impact non-original films distributed in theaters. The eWOM for non-original films released in theaters will experience the effects of $\beta_1 + \beta_2 + \beta_3$. eWOM for original films released in theaters will experience the effects of β_2 . eWOM for non-original films released through streaming services will experience the effects of β_1 . eWOM for original films released through streaming services will experience none of the effects of the first three betas.

A similar group of controls to those present in Test Three will also be necessary for the experiment. Controlling for year is important given the short time frame during which streaming-based original series and films have been available, meaning the first years of these innovations could experience different effects on eWOM than in later years as consumers became accustomed to the new developments. Like Experiment 2, the analysis for Experiment 3 will also control for profile, critical ratings, genre, and release month (MPAA was unavailable on Netflix during the time period of the research).

How does eWOM translate to sales?

Within the dataset, the indices for eWOM during Week 1 and 2 are 83.8 and 83.6 percent correlated with the total real revenue of the film, respectively. This provides an initial indication that eWOM from the two opening weeks can serve as a reliable proxy for total sales. The two indices themselves are 94.7 percent correlated, leading both indices to be almost equally correlated with revenue. Both indices are less correlated with opening weekend gross, with only 81.7 and 78.6 percent correlation for eWOM during Week 1 and 2 and opening weekend revenue, respectively. This drop may be due to moviegoers demonstrating interest in a film by searching it online or reading reviews during Week 1 or 2 but planning to see it at a later date. In this case, their interest would only affect the total revenue, rather than the opening weekend revenue. Thus, Test 4 will consider total box office as sales.

A simple regression of real revenue on both indices for Week 1 and Week 2 can show an

	Dependent variable:	
	Real Rev (MM USD)	
	(1)	(2)
Ind.1	16.847*** (0.595)	
Ind.2		12.242*** (0.435)
Constant	35.227*** (4.866)	39.901*** (4.787)
Observations	341	341
R ²	0.703	0.700
Adjusted R ²	0.702	0.700
Residual Std. Error (df = 339)	70.892	71.160
F Statistic (df = 1; 339)	801.340***	792.746***

Note: *p<0.1; **p<0.05; ***p<0.01

T11: Real Revenue in millions of USD regressed on the two weekly indices for eWOM. Coefficients represent millions of 2018 USD.

approximate marginal effect of eWOM on sales. For each additional unit increase in the index for Week 1, real revenue is expected to increase by \$16.8 million dollars. Likewise, if the second index was instead used to predict sales, for each additional unit increase in the index for Week 2, real revenue is expected to increase \$12.4 million. To contextualize this difference, *Avengers: Infinity War* is four units higher than *Jurassic World* on the index for Week 1, and five units higher on the index for Week 2. This would translate into expecting *Infinity War* to earn about

\$58 million or \$60 million more than *Jurassic World*, depending on the index used, over its entire box office run. While this estimation does not properly reflect the differences in financial success between these two specific films, this example does demonstrate the sizable expected financial effect of unit changes in eWOM within the regression.

Results of Primary Test Four Regression

	<i>Dependent variable:</i>				<i>Dependent variable:</i>		
	eWOM Index - Week 1				eWOM Index - Week 2		
	(1)	(2)	(3)		(1)	(2)	(3)
Unor	0.038 (1.311)	0.045 (1.313)	-0.007 (1.172)	Unor	0.014 (1.817)	0.029 (1.819)	0.099 (1.573)
Theater	2.717*** (0.968)	2.865*** (1.017)	-2.242** (1.073)	Theater	3.297** (1.328)	3.526** (1.394)	-3.782*** (1.439)
Unor:Theater	2.688* (1.472)	2.659* (1.475)	1.428 (1.322)	Unor:Theater	3.915* (2.036)	3.866* (2.040)	1.950 (1.773)
Time		0.106 (0.220)	-0.176 (0.197)	Time		0.164 (0.302)	-0.253 (0.264)
Action			1.871** (0.753)	Action			2.387** (1.010)
Adventure			1.740*** (0.665)	Adventure			2.776*** (0.892)
Horror			2.275** (0.992)	Horror			2.864** (1.331)
RT.Score			-0.017 (0.011)	RT.Score			-0.011 (0.015)
RT.NumRevs			0.031*** (0.004)	RT.NumRevs			0.044*** (0.005)
Apr			1.635 (1.101)	Apr			1.901 (1.477)
Jun			2.337*** (0.880)	Jun			2.434** (1.180)
Dec			1.694** (0.812)	Dec			2.222** (1.090)
Constant	0.428 (0.793)	-1.409 (3.904)	2.191 (3.529)	Constant	0.526 (1.087)	-2.330 (5.354)	2.406 (4.733)
Observations	423	423	422	Observations	422	422	422
R ²	0.123	0.124	0.335	R ²	0.119	0.120	0.359
Adjusted R ²	0.117	0.116	0.315	Adjusted R ²	0.113	0.112	0.340
Residual Std. Error	5.720 (df= 419)	5.725 (df= 418)	5.043 (df= 409)	Residual Std. Error	7.842 (df= 418)	7.848 (df= 417)	6.766 (df= 409)
F Statistic	19.667*** (df= 3; 419)	14.781*** (df= 4; 418)	17.154*** (df= 12; 409)	F Statistic	18.886*** (df= 3; 418)	14.215*** (df= 4; 417)	19.058*** (df= 12; 409)
<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01	<i>Note:</i>			*p<0.1; **p<0.05; ***p<0.01

T12 (Left) and T13 (Right): In both cases, the eWOM index for Week 1 (left) and Week 2 (right) is regressed on binary variables for originality and distribution model, as well as control variables similar to those used in Test Three. Coefficients are of the magnitude of the unitless indices.

Discussion

In both the first and second regression of T12 and T13, the impact of the traditional film release on the eWOM of the film immediately stands out. For both original and non-original films, the eWOM is expected to rise by sizable and statistically significant amounts. For example, a film's Week 1 eWOM is expected to increase by 2.717 points if released through a theater, and an additional 2.688 points if it is also non-original. A similar trend is evident for Week 2 eWOM, with expected increases of 3.297 if released in theaters and an additional 3.915

if also original. Much like Test 3, when also controlling for time, no statistically significant trend emerges, as seen in the second regression. Without the control variable for film profile, which is only featured in the third regression in both tables, this result seems plausible. Most of the Netflix Original films receive less attention than movies which were among the highest-grossers at the domestic box office and would thus generate less eWOM than their traditional counterparts. Therefore, the coefficients for the theater variable—economically and statistically significant—demonstrate the need for further controls (primarily the film profile).

With the inclusion of all relevant control variables (non-relevant controls have already been removed) in the third regression of T12 and T13, the coefficient for Theater; (β_2) reverses its sign while maintaining its statistical significance. This proves that while holding the profile of a film constant, there is a negative effect felt by original films released in theaters that is not cancelled out by any potential positive increase from the coefficient for non-originality (β_1) or the interaction term (β_3). To compare between original and non-original theatrical films, there is a potential for original films to generate $\beta_1 + \beta_3$ less eWOM than non-original films. eWOM is expected to be 2.24 and 3.782 points lower in Weeks 1 and 2, respectively, for theatrical films compared to Netflix films holding originality constant. To contextualize this in financial terms using the regressions in the previous subsection, this would lead to an approximate decrease of \$37.6 million and \$46.3 million in revenue based on the eWOM regressions for Week 1 and Week 2, a strong economic effect.

In similar fashion to Test Three, the coefficient for the interaction term (β_3) in the third regression in both T12 and T13, representing non-original films released in theaters, shows a sizable positive but statistically insignificant expected increase in eWOM. Like the interaction term between non-originality and time period discussed earlier in Test Three, the lack of

significance may indicate other control variables better capture the impact of non-originality. To reiterate, summer and holiday season releases, as well as the action and adventure genre, are often present for superhero or franchise films, which may capture the effect of non-originality more than the binary variable does alone.

Looking at the effects only felt by streaming-based content, non-original films released through streaming services are expected to see no increase in eWOM relative to original films in both T12 and T13. The absence of a significant or sizable coefficient (β_1) over this term works to partially prove the second hypothesis—that non-original content does not enjoy the same competitive advantage in the new online streaming space—by showing there is no difference in eWOM between original and non-original Netflix titles.

Thus, acknowledging the lack of significance for β_1 and β_3 , a conservative interpretation of the data suggests films released in theaters, holding all other variables constant, perform worse than do Netflix releases. However, like the discussion in Test 3, much of the effects which could be represented by β_3 are likely captured by other control variables—specifically the action and adventure genres, summer and holiday releases, and large profiles—due to the frequent cross-over between non-originality and these variables. If the sizable effect of β_3 was acknowledged, despite its lack of significance, then non-original theatrical releases would receive 1.428 and 1.950 greater eWOM in Weeks 1 and 2, respectively, than would original films. On the other hand, as established earlier, β_1 is neither sizable nor significant. Thus, the marginal effect felt by non-original versus original films for theatrical releases, $\beta_1 + \beta_3$, would be greater than the marginal effect felt by non-original versus original films for Netflix release, β_1 . This scenario, although imperfect, presents the strongest case for the validity of the paper's second hypothesis.

Test Four requires further investigation and expansion in order to properly validate the second hypothesis. Acquiring more data points by utilizing more films, both released in theaters and across Netflix and other distributors entering the space, will quickly raise the economic significance of many of the variables. Additionally, the employment of a more efficient tool to collect the eWOM of the various individual films will allow for the elimination of biases related to common phrase titles and total weekly eWOM counts, as well as provide a better method for reducing endogeneity due to eWOM's status as both preceding and following sales. With a stronger experimental design and the release of more streaming-based films, the question of whether original content can succeed relatively more on online-streaming platforms can be answered more precisely in the future.

VI. Conclusion

The research present in this paper set out to prove two hypotheses; first, the success of non-original content has increased, especially in recent years, within the traditional distribution model; second, original content would fare better compared to non-original content in the new streaming platform space relative to its success in the traditional distribution model. The results offered in Tests One, Two, and Three strongly support the first hypothesis, while Test Four suggests the potential validity of the second hypothesis. Likewise, the findings of this paper provide several implications for the future of filmmaking.

For movie studio executives deciding how to assemble their portfolio of theatrical releases in the coming years, investing heavily in known properties rather than attempting to strike gold with original ideas (after all, "Nobody knows anything") remains a sound financial option. As demonstrated by the tests present throughout the paper, non-original films dominate

Hollywood, specifically comic book and sequel projects. While this was never always the case, the stranglehold held by such franchise films over the box office has only tightened in recent years, in no small part due to the changes in entertainment options available to consumers on their laptops, smartphones, and televisions.

For aspiring filmmakers who hope to create innovative content of their own, the opportunities provided by online distribution platforms may actually open the door for original films to succeed like never before. In coming years, the “golden tail of digitization” already being experienced by television shows enjoying audiences more receptive to unfamiliar content may soon apply to films produced specifically for distribution online. The prospect is exciting for filmmakers hoping to create high-quality original content which can be appreciated by audiences previously unavailable in theaters. Only this year, *Roma*, a Netflix Original title, received an Oscar Nomination for Best Picture. Perhaps a sign of things to come, streaming services may soon take over the artistic film-space, much like they already have with television.

However, given the extremely recent introduction of Original films, further exploration of this topic is needed before the success of original films on streaming platforms can be proven. The film industry is a constantly changing space, and a four-year period of study is simply not long enough to assume original films will succeed on online platforms in the distant and even intermediate future of filmmaking.

Works Cited

- Baker, W., Hutchinson, J., Moore, D., Nedungadi, P. (1986). Brand Familiarity and Advertising: Effects on the Evoked Set and Brand Preference, *Advances in Consumer Research*, 13: 637-642.
- Berlyne, D.E. (1970). Novelty, Complexity, and Hedonic Value, *Perception & Psychophysics* 8: 279. Retrieved from: <https://doi.org/10.3758/BF03212593>
- Box Office Mojo. Retrieved from www.boxofficemojo.com/.
- D., W., G., B., W., & B., A. (2008, March 04). The Dynamics of Online Word-of-Mouth and Product Sales: An Empirical Investigation of the Movie Industry. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0022435908000171>
- Dhar, T., Sun, G. & Weinberg, C.B. Mark Lett (2012). The Long-term Box Office Performance of Sequels, *Marketing Letters*, 23: 13. Retrieved from: <https://doi.org/10.1007/s11002-011-9146-1>
- Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4), 545+. Retrieved from http://link.galegroup.com.proxy.lib.duke.edu/apps/doc/A126388226/ITOF?u=duke_perkins&sid=ITOF&xid=e6cc1b34
- Hiller, R. S. (2017). Profitably Bundling Information Goods: Evidence from the Evolving Video Library of Netflix. *Journal Of Media Economics*, 30(2), 65-81. doi:<http://dx.doi.org.proxy.lib.duke.edu/10.1080/08997764.2017.1375507>
- Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2. <https://CRAN.R-project.org/package=stargazer>
- Katz, E., & Lazarsfeld, P.F. (1955). *Personal Influence*. Free Press, Glencoe, IL.
- Koblin, J. (2017, October 18). How Many People Watch Netflix? Nielsen Tries to Solve a Mystery. Retrieved from <https://www.nytimes.com/2017/10/18/business/media/nielsen-netflix-viewers.html>
- Lehrer, S., & Xie, T. (2017, December). Box Office Buzz: Does Social Media Data Steal the Show from Model Uncertainty When Forecasting for Hollywood? Retrieved from https://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00671
- Richwine, L. (2018, November 01). Netflix to release 3 films in theaters ahead of online debut. Retrieved from <https://www.reuters.com/article/netflix-movies/corrected-netflix-to-release-3-films-in-theaters-ahead-of-online-debut-idUSL2N1XB2J6>

- Rodriguez, A. (2018, April 14). *Twenty Years Ago, Netflix.com Launched. The Movie Business Has Never Been the Same*. Retrieved from qz.com/1245933/twenty-years-ago-netflix-com-launched-the-movie-business-has-never-been-the-same/.
- Slack, M. (1999). *Guerilla marketing breaking through the clutter with word-of-mouth*, Jupiter Research.
- Spence, A., (1976), Product Differentiation and Welfare, *American Economic Review*, 66, issue 2, p. 407-14, <https://EconPapers.repec.org/RePEc:aea:aecrev:v:66:y:1976:i:2:p:407-14>.
- TheNumbers. Retrieved from <https://www.the-numbers.com/>.
- Waldfogel, J. (2017). The Random Long Tail and the Golden Age of Television. *Innovation Policy and the Economy*, 1-25.
- Wolff-Mann, E. (2018, January 09). Google Trends is a surprisingly good way to predict Netflix subscriber growth. Retrieved from <https://finance.yahoo.com/news/google-trends-surprisingly-good-way-predict-netflix-subscriber-growth-142200691.html>