What Took So Long? An Analysis of Survival Times for Movie Sequels

Alexander Hakemoller Marsh and Colin Cannonier*

ABSTRACT. Movie sequel windowing is one area that has been neglected in the existing economic literature. While there are many potential areas for economic research in the film industry, the lack of research specifically on sequels and their timing is rather surprising amidst the increasing saturation of sequels in the market place. Using a dataset constructed from an IMDb list and data from The Numbers, this research models the timing between sequels using an accelerated failure time survival model in order to understand when film studios decide to release a sequel. The results show that while film studios do release sequels quicker the more money the previous movie makes, the decrease in time is inelastic and not economically significant. The most significant factor that influenced timing between sequels was whether the movie was produced by the same studio as the previous movie. Time between sequels is increased 69% (752 days) whenever a different studio is producing the sequel. Similarly, time between sequels decreases by 45% (487 days) if the previous movie in the series was profitable, by 44% (482 days) if produced by a major film studio, and by 29% (319 days) if the movie or series is based on a preexisting work. The results are robust to a variety of sensitivity checks related to holiday releases, ratings, genres and period-specific effects. (L82, Z11)

I. Introduction

Movie sequels have become increasingly popular in the film industry in recent years. Film studios see sequels as a version of brand extension and a safer choice for turning a profit (Basuroy and Chatterjee 2008). As the share of sequels in the top grossing films continues to rise, econometric research directed specifically at sequels becomes ever more valuable and important. To the economist, how film studios determine the length of time between sequels is particularly interesting since varying this time is one way to control the supply of their films (De Vany and Walls 1997). Even though there are researchers, including economists, focusing their attention on the film industry (De Vany and

---

*Colin Cannonier, Corresponding author. Jack C. Massey College of Business, Belmont University, 1900 Belmont Boulevard, Nashville, Tennessee, USA 37212. Email: colin.cannonier@belmont.edu Thanks to participants at the Missouri Valley Economic Association (MVEA) 53rd Annual Conference for providing helpful comments and suggestions. We are grateful to the editor and two anonymous referees whose suggestions improved the paper substantially.

Walls, 1997; Walls, 2009; Einav, 2002; Bohnenkamp, Knapp, Hennig-Thurau, and Schauerte, 2015), the amount of research directed at such a mature field is surprisingly sparse.

While there is marketing and generic business research directed at the film industry, these studies infrequently use sophisticated econometric techniques to estimate parameters and test hypotheses. Furthermore, the amount of research directed specifically at sequels is even more limited. Of the existing literature, Basuroy and Chatterjee's (2008) research is the closest to a comprehensive analysis of movie sequels and their profitability. The authors conducted a thorough analysis of movie sequel profitability by testing a handful of hypotheses related to the box office performance of movie sequels. Using weekly box office data and the Generalized Estimating Equations method, they examined the relationship between a sequel's revenue and the following factors: revenue of the parent film, the amount of time between the release of the sequel and the release of the parent film, the number of pre-existing movies in the series, the revenue of non-sequels, and the difference in the rate of change of weekly revenue for sequels and non-sequels. They found that sequels do not match the box office performance of the parent film; however, they do perform better than non-sequel films. Their results also suggest that sequels perform best when they are released closer to the parent film rather than later, which is particularly interesting to this research. Our work expands upon theirs by using a larger sample size (theirs used 167 observations) extending to more recent data (theirs covered the period 1991 to 1993). If this relation still holds, it would imply that a movie studio should minimize factors that lengthen the time between sequels in order to maximize revenue.

De Vany and Walls (1997) have conducted some of the most notable econometric research focused on the film industry. In their paper, they used survival analysis to determine the factors that affect how long movies stay in theaters. They found that the rate at which movies leaves theaters (i.e. the hazard rate) is not constant and increases with time. The most significant explanatory variable for increasing survival time was the number of theater bookings when the movie is first released. While this paper isolated the important variables for increasing survival time and has proven valuable to understanding the competitive nature of box office survival, a similar analysis has not been conducted for movie sequels and their release times. Walls (2009) also conducted a nonparametric analysis to examine movie profitability through many lenses, one of which was the sequel. Consistent with the findings by Basuroy and Chatterjee's (2008), he found that sequels were more profitable than non-sequels. Considering the increasing importance of sequels to the film industry, our research intends to extend the literature and help us better understand what factors influence time between sequels. This is important as it will help aid future research into optimal sequel timing windows.

In this paper, we hope to contribute to the literature by determining the factors that influence the time between sequel releases by film studios. While there is research into the game theoretical nature of movie releases in general (Einav 2002), we are unaware of any significant work specifically focusing on sequels. From an economic perspective, understanding the factors associated with the release of sequels is important because time between sequels is expected to influence other important factors such as box office sales, demand, survivability of the series, and consumer perception. If timing is too quick, then consumers' preferences may be saturated leading to underperformance. On the other hand, if timing is too late, consumers' tastes might change, also leading to underperformance. Rational studio operators will aim to minimize delays (or time to release) so that an optimal release window is achieved that maximizes sales. Accordingly, this research models time between releases mainly through the lens of monetary influences (example: sales of the previous film, the budget of the sequel, and whether the previous movie was profitable) and studio influences such as whether the sequel or series is based on a preexisting universe and whether the sequel is being produced by the same studio as the previous film. Other factors such as genre and maturity rating are considered.

Using a data set of 674 films with 406 sequels, the time between sequels is modeled using an accelerated failure time survival model. The rest of the paper proceeds as follows: Section II provides a brief overview of survival analysis; Section III outlines the estimation technique and model selection; Section IV discusses the data, variables, and sample; Section IV provides an analysis of the results and Section VI concludes.

II. Survival Analysis and Movie Sequels

Modeling time to a sequel in the film industry is important because adjusting the length of this gap is one way in which film studios try to adjust the supply of their products in order to meet demand so that profits are maximized (De Vany and Walls 1997). As a result, understanding what factors affect the time between sequels can help studios make rational releasing decisions. Since time between sequels can be modeled using "birth" or "death" process, survival analysis can be utilized in order to understand sequel timing. Birth is defined as whenever the series is "at risk" of acquiring another sequel. For the first sequel in a series, this is whenever the parent film (the "zeroth" sequel) is released; for the second sequel, this is whenever the first sequel is released, and so forth. At this moment, time is equal to zero and the series is at risk of getting a new sequel. Death is defined as whenever the sequel in question is released and the clock stops. Therefore, the survival time is the time from birth to death and is denoted as the random variable, T, with a probability density function (pdf) and a cumulative density function (cdf). Because the cdf returns the probability that the event T is less than or equal to time t, the complement of the cdf returns the probability that event T is greater than t, or in other words, the probability that the respondent is still alive at time t (Rodriguez 2007). This is referred to as the survival function, which is the foundation of survival models.

Following from the above, this research utilizes a model derived from the hazard function or the hazard rate. The hazard rate is defined as the instantaneous rate of occurrence at T=t, or the probability that T occurs in a certain interval given that T has yet to occur at the start of some defined interval. It can be shown that the hazard rate is the probability that T occurs in some interval defined by the pdf divided by the likelihood that T has not already occurred.

III. Estimation Technique and Model Selection

To model time between sequels, an accelerated failure time (AFT) model was utilized. Let the variable time, t, be expressed as follows: $t = e^{x'\beta}v$ where x is a vector of covariates, v=e^u and u follows a specified probability distribution. Then the hazard rate is defined as $\lambda(t|\mathbf{x}) = \lambda_0(v)e^{-x'\beta}$, where $\lambda_0(v)$ is the baseline hazard which does not depend on t (Cameron and Trivedi 2005). Since $v = te^{-x'\beta}$, by substitution, the hazard rate can be expressed as an explicit function of t,

$$\lambda(t|\mathbf{x}) = e^{-x'\beta}\lambda_0(te^{-x'\beta}). \tag{1}$$

If $e^{-x'\beta} > 1$, then the baseline hazard is accelerated which shortens the survival time. Likewise, if $e^{-x'\beta} < 1$, then the baseline hazard is decelerated, leading to a lengthening of survival time (Cameron and Trivedi 2005). Now, let T be the random variable survival time with an unspecified probability distribution. Using log properties, log T can be written as $\log(T\, e^{-x'\beta}) - \log e^{-x'\beta}$, which simplifies to $\log(Te^{-x'\beta}) + \boldsymbol{x}'\beta$. From above, $v = T\, e^{-x'\beta}$, therefore $\log(T\, e^{-x'\beta})$ represents the error term, u. If $\log T e^{-x'\beta}$ is distributed independently of $\boldsymbol{x}'\beta$, then we can write the log of survival time as,

$$\log T_i = x_i'\beta + u_i, \tag{2}$$

where x_i is a vector of covariates of the i^{th} observation and u_i is the residual of the i^{th} observation, whose distribution is dependent upon the selected distribution of T.

Note that this is very similar to a standard OLS log-linear regression with log of time as the dependent variable. The difference between the two techniques is that AFT models are estimated using maximum likelihood, they allow for other distributions of T and u beyond the standard normality assumptions in classical linear regression and such modes allow for censored data in their maximum likelihood estimation. Inserting the vector of variables (discussed in Section V) into equation 2 gives the following model:

$$\log(T_{ni}) = a + \beta Monetary'_{ni} + \eta Production'_{ni} + \delta Control'_{ni} + \mu_{ni}, \tag{3}$$

where a is the constant representing the average time to sequel, $Monetary_{ni}$ is a vector of monetary variables for sequel n of series i, $Production_{ni}$ is a vector of production variables for sequel n of series i, $Control_{ni}$ is a vector of control variables as well as sequel dummies n of series i, and μ_{ni} is the error term. The above regression will be estimated using a maximum likelihood estimation technique which is standard in survival analysis.

To control for the fact that different sequel numbers, might have different distributional parameters, the data will be stratified across sequel numbers so that a separate scale parameter, σ_n , for each sequel number will be estimated. The dependent variable, T, is assumed to have a log-normal distribution. As such, the probability density function is as follows,

$$f(x) = \frac{1}{x\sigma_n\sqrt{2\pi}} e^{\frac{\ln x - a^2}{2\sigma_n^2}}, \tag{4}$$

where a is the location parameter which represents the average time between sequels estimated in equation 3 and σ_n is the scale parameter of sequel n. The associated cdf, survival function, and hazard function can be derived from equation 4.

IV. Data

The data for this research were primarily collected from two sources: *The Numbers*, Nash Information Services, LLC and the *Internet Movie Database* (IMDb). Table 1 reports the descriptive statistics of the data for each of the main variables discussed. Most of the data including sales revenues, budgets and production studios were taken from *The Numbers*. *The Numbers* is one of the leading providers of movie industry data and serves around 1,000 clients in various parts of the industry. As such, the quality of the data is high. IMDb was used to corroborate the data from *The Numbers* and fill in any missing data. Occasionally, we supplement movie budget data for select movies with budget estimates from IMDb. We also provide the correlations between these variables in Appendix Table 2.

The dependent variable time will be considered a continuous variable in this research. Time is being measured in days and therefore has the smallest discrete interval possible, given how movies are released.

TABLE 1–Descriptive Statistics

	Observations	Median	Mean	Standard Deviation	Minimum	Maximum
Time	406	1,092	1,612	1,577.0	174	10,900
Sales.of.Prev.Movie	406	$64.0	$80.9	$80.3	$0.02	$822.8
Budget	406	$26.3	$35.2	$29.4	$0.82	$146.1
Profitable.Prev.Movie	406	1	0.92	0.27	0	1
Profit.of.Prev.Movie	406	$34.5	$51.1	$74.1	-$38.5	$748.8
PreExisting.Universe	406	0	0.35	0.48	0	1
Different.Studio	406	0	0.16	0.36	0	1
Major.Studio	406	1	0.77	0.42	0	0
Sequel Number	406	1	1.80	1.09	1	5
G-rated	406	0	0.03	0.18	0	1
PG-rated	406	0	0.22	0.41	0	1
PG 13-rated	406	0	0.38	0.49	0	1
R-rated	406	0	0.36	0.48	0	1
NC17-rated	406	0	0.007	0.086	0	1
Year.of.Prev.Movie	406	2001	1998	11.65	1960	2015
Month	406	5	5.60	3.09	1	12
Prequel	406	0	0.08	0.27	0	1
Horror	406	0	0.24	0.43	0	1
Action	406	0	0.41	0.49	0	1
Thriller	406	0	0.22	0.41	0	1
Adventure	406	0	0.36	0.48	0	1
Comedy	406	0	0.31	0.46	0	1
Family	406	0	0.09	0.28	0	1
Animation	406	0	0.06	0.25	0	1
Superhero	406	0	0.06	0.24	0	1
SciFi	406	0	0.17	0.38	0	1
Fantasy	406	0	0.14	0.35	0	1
Drama	406	0	0.12	0.33	0	1
Mystery	406	0	0.11	0.32	0	1
Romance	406	0	0.05	0.21	0	1
Music	406	0	0.01	0.11	0	1
Sport	406	0	0.02	0.13	0	1
Crime	406	0	0.14	0.35	0	1
Biography	406	0	0.002	0.05	0	1
History	406	0	0.005	0.07	0	1

Furthermore, whenever the discrete intervals are relatively small compared to the average time to event, the data can be considered continuous (Allison 1982). Although the discrete interval of this data is one day, the median time to event is 1092 days or approximately 3 years. Continuity is therefore assumed.

A. SAMPLE

The movies considered for this research were selected from an *Internet Movie Database* (IMDb) user-generated list of all sequels (Kreuzberg 2011). It is the most comprehensive list of movie sequels found thus far. In order for a movie to be included in the dataset, it had to have a U.S. theatrical release and be a member of a series that had the original film and at least one sequel. The original film that starts the series will be referred to as the parent film. Since the topic of this research is modeling time between sequels, the parent films were filtered out of the dataset so only the effects of the covariates on sequels would be estimated. Films that are considered "reboots" of an existing series (e.g. *Spider-man* vs. *The Amazing Spider-man*) were treated as a new parent film and the start of a new series. This was to help control for series/casting continuity. Lastly, because having six or more sequels is very rare, a subset of the dataset was used that excluded any sequel six or greater. After excluding all incomplete data, we were left with 406 observations.

B. VARIABLES

The dependent variable is the log of time in days between the sequel (the n^{th} sequel in series i) and the previous movie (the n-1 sequel in series i). The key variables in the model comprise of the following inputs that are considered important in determining the timing of a sequel release. The monetary inputs consist of sales of the previous movie (*Sales.of.Prev.Movie*), *Budget*, and a categorical variable for whether the previous movie in the series was profitable (*Profitable.Prev.Movie*). Due to better goodness-of-fit measures and higher collinearity issues with the continuous profit variable (see correlation matrix in Appendix Table 2), our main specification uses the categorical form. As a robustness check, we show that the results are largely unchanged when profit of the previous movie is measured as a continuous variable.

Sales.of.Prev.Movie is defined as the domestic box office revenue that the previous movie in the series earned and is measured in millions of U.S. dollars adjusted for inflation. The sign for the *Sales.of.Prev.Movie* coefficient is expected to be negative. If the sales of the previous movie increase, then a studio will want to release a sequel as quickly as possible in order to capitalize on the success of the film, thus decreasing the time between sequels. *Budget* is the amount budgeted for the film's production and is measured in millions of U.S. dollars adjusted for inflation. The coefficient of this variable is expected to be negative as well. If a film has a larger budget, the studio should be able to produce that film quicker. Similarly, *Profitable.Prev.Movie*, defined as sales minus production budget, is expected to carry a negative coefficient. While there are other expenses associated with a movie's production and distribution (advertising, marketing, distribution, etc.), profit is defined as such because of data limitations and the notorious accounting practices in Hollywood (Daniels, Leedy, and Sills 1998). First, *The Numbers* only reports the production budget of films in their database. Notwithstanding the data limitations, film studios are known for "creative" accounting practices such as reporting the widest array of expenses possible under the highest grossing film so that loses are shared. This makes any analysis of profit beyond sales minus production budget less informative. Furthermore, studios will be most concerned about whether a movie justified its production costs, as production is usually the largest expense.

The production variables consist of a set of categorical variables that are measured as follows: whether the characters, universe, or plot lines of the sequel or series were based on a preexisting work or historical event (*PreExisting.Universe*); whether the sequel was produced by a different studio than the previous film in the series (*Different.Studio*); and whether the movie was produced by one of the six major film studios (*Major.Studio*). *PreExisting.Universe* if a dichotomous variable that takes the value of one if sequel is an extensuion of another universe and zero otherwise. When a movie is based on a previous work, it should decrease the amount of mental and physical labor required to produce a movie, thus it should theoretically decrease the time between sequels. Therefore, the expected sign for *PreExisting.Universe* should be negative. *Different.Studio* is a dichotomous variable that takes the value of one if the sequel is being produced by a different studio and zero if the same studio is producing the sequel. The reasoning for this variable is that a studio can either merge with other studios or buy the rights of a film series from another studio. Whenever this happens, planned movie sequels can be delayed as the rights are moving from one entity to another. Similarly, sometimes these moves can resurrect failed series that have not seen a sequel for some time. Consequently, the expected sign for *Different.Studio* is positive since if a movie goes from zero (same studio) to one (different studio), one would expect the time between sequels to increase. Lastly, *Major.Studio* is a dichotomous variable that takes on the value of one if the movie was produced by one of the six major film studios, Warner Bros. Pictures, Universal Pictures, Walt Disney Pictures, 20th Century Fox, Columbia Pictures (Sony), and Paramount Picture, and zero otherwise. The expected sign for *Major.Studio* is negative because major studios should have more resources and connections which lead to quicker movie production. The model controls for genre, maturity rating assigned to the movie as well as year dummies. Appendix Table 1 provides details on these variables.

V. Results

Our regressions were run in R using the "survival" package and the *survreg* function (Therneau 2015). Table 2 presents the results of equation 3 including controls sequentially by column. *Sales.of.Prev.Movie* and *Budget* were logarithmically transformed to make the variables more linear and to allow the coefficients to have an elasticity interpretation. Column 1 presents the results of the model comprising only monetary variables. In column 2, we add production variables and thereafter we control for year effects (column 3), then for genre and rating effects (column 4), before stratifying by sequel number (column 5). All specifications are estimated with robust standard errors. The strata function from the R "survival" package was used to run each sequel number as a different stratum of data. The results are robust to various model specifications; that is with the inclusion of other control variables and dummy variables. Our preferred specification with the most explanatory power is obtained from the results in column 5. Accordingly, interpretation and discussion of the parameters will be based on these estimates.

TABLE 2–Estimates from Log-Normal Regression: Time-to-Sequel Release in Days

Variable	Dependent Variable: log(Time)				
	(1)	(2)	(3)	(4)	(5)
Intercept	7.433*** (0.182)	6.995*** (0.169)	8.996*** (0.212)	9.144*** (0.396)	9.158*** (0.392)
log(Sales.of.Prev.Movie)	-0.153*** (0.053)	-0.129*** (0.047)	-0.164*** (0.046)	-0.159*** (0.043)	-0.148*** (0.043)
log(Budget)	0.205*** (0.051)	0.237*** (0.046)	0.256*** (0.048)	0.226*** (0.055)	0.217*** (0.058)
Profitable.Prev.Movie	-0.444*** (0.169)	-0.466*** (0.162)	-0.513*** (0.150)	-0.540*** (0.137)	-0.591*** (0.128)
Different.Studio		0.630*** (0.108)	0.541*** (0.098)	0.542*** (0.096)	0.524*** (0.090)
Major.Studio		0.415*** (0.081)	0.380** (0.080)	0.354*** (0.085)	0.366*** (0.085)
PreExisting.Universe		-0.435*** (0.072)	-0.357** (0.069)	-0.339*** (0.085)	-0.346*** (0.081)
Year controls	No	No	Yes	Yes	Yes
Genre controls	No	No	No	Yes	Yes
Rating controls	No	No	No	Yes	Yes
n-Stratification	No	No	No	No	Yes
Observations	406	406	406	406	406
Pseudo R^2	0.098	0.272	0.426	0.479	0.482
Chi-Squared	41.92***	129.06***	225.20***	264.60***	266.81***
Scale (Sequel 1)	0.738***	0.663***	0.589***	0.561***	0.516***
Scale Sequel 2					0.551***
Scale Sequel 3					0.785**
Scale Sequel 4					0.607**
Scale Sequel 5					0.432*

Notes: Robust Standard errors are in parentheses below the coefficients. Statistical levels of significance are: *indicates significant at the 10 percent level; ** indicates significant at the 5 percent level; *** indicates significant at the 1 percent level. Pseudo R^2 were calculating using the formula, $R^2 = 1 - e - \frac{LRT}{n}$, where LRT = −2logL(null)−(−2logL(fitted)), and n is the sample size (Allison 1995).

All coefficients have their hypothesized sign except for *Budget* and *Major.Studio,* which were both hypothesized to be negative. Importantly, however, these two variables have the same sign. Movies produced by major studios typically have a larger budget (median budget: $31.9 million for majors and $13.6 million for non-majors). Likewise, films with extravagant special effects and large scale production have bigger budgets, and consequently, take longer to produce because of the scale of the production. It is also possible that if a film studio is putting forth more money, there is a greater incentive to make the best possible movie so that the sequel is profitable. Major studios have more capital invested and higher opportunity costs associated with an underperforming film. In these circumstances, production may be delayed in order to guarantee a successful film. Time between sequels is about 44% (about 482 days, or about 1.32 years) longer, on average, if the film is produced by a major studio. However, while it does appear that larger budgets increase time between sequels, the relationship is not very strong. Since this coefficient can be interpreted as an elasticity, it appears that studios are not less sensitive to changes in the budget. Specifically, an increase of the budget by 25% would only increase time to sequel by 5.4% (or 60 days on average).

It appears that the timing of the release of movies by studios is also less sensitive to changes in the sales of the previous movie. A 1% increase in the sales of the previous movie only leads to a 0.15% decrease in time to sequel (about 2 days). This may appear counterintuitive at first. Would studios not want to release a sequel as soon as possible if the parent film massively outperforms sales? This result can be reconciled when considering the limiting factors influencing film production. Studios are limited by the necessary time needed to produce a film. While technology has somewhat alleviated this constraint, there are still few movies that can be produced in less than a year. Consequently, even if sales of the previous movie are high, a studio is unable to turn around and immediately release a sequel. Secondly, the scheduling process for each series is vastly different. While the releases of some sequels within a series are planned for discrete intervals far in advance (*Lord of the Rings, Twilight, The Hunger Games*, etc.), others seem to be stochastically selected for continuation by a studio. These series-specific factors are probably more influential when determining time between sequels than any of the quantifiable, explanatory variables included in this research. While the revenue of a film may be influential in determining whether a sequel is produced, it does not affect the time between sequels. Lastly, it is well known that there is much strategic behavior involved in deciding when to release movies in general. Not only do studios account for the time to produce the film, they also have to consider factors such as what other movies will be in theaters at the same time, seasonal demand, and future consumer tastes. Based on the regression coefficients, this appears to be true for sequels as well. If sales are not an *economically* significant factor, then these strategic decisions may be more important when deciding when to release a sequel. Such strategic factors will also be discussed in further detail in the robustness checks.

The remaining three coefficients have a fairly intuitive interpretation. If a sequel is being produced by a different studio than the previous film, time between sequels increases by about 69% (i.e. $100[e^{0.524} - 1]$) relative to those that were produced by the same studio. Similarly, if the previous movie in the series was profitable, time between sequels decreases by about 59% (or a year and 9 months). Lastly, if the sequel or series is based on a preexisting universe, time between sequels decreases by 35% (382 days, just over a year). While the direction of these changes is rather intuitive, their magnitudes are rather large. In fact, these appear to be the more economically significant coefficients. Simply by changing production studios resulted in a 69% delay (753 days or just over 2 years) in release time. Given that the median time between sequels for this dataset is about 1092 days, a 60% increase in the time is significant.

A. ROBUSTNESS CHECKS

To investigate the sensitivity of the results, we conduct several robustness checks. These results are shown in Table 3 (columns 2 to 7). For the ease of comparison, the results from the preferred specification are reported in column 1. The first sensitivity analysis involves analyzing the effect of profit of previous movie as a continuous variable rather than a dichotomous measure. The estimates from this measure are less informative since a dichotomous measure likely reduces its usefulness by treating marginally profitable films in the same way as highly profitable ones. Since profit can be negative, we undertake the following log-modulus transformation: the transformation: $L(x) = sign(x) * \log(|x|)$. Column 2 provides the results of this transformation. Similar to the main specification, the coefficient on the continuous measure

Proft.of.Prev.Movie is also negative and statistically significant. All other coefficients carry the same sign and remain statistically significant. It is useful to note that the preferred specification has stronger explanatory power when profit of previous movies is used a dichotomous variable.

Next, we consider some other issues regarding the sequel. Specifically, we include a dichotomous measure called "prequel" which takes the value of one if the sample include prequels and zero otherwise. The results of the inclusion of this variable in our main specification are shown in column 3 of Table 3. We find the coefficient to be insignificant. This insignificant finding is consistent with the general observation that the chronological order of the story tends to have little effect on the timing of a movie release. In fact, based on our sample, prequels account for less than 10 percent (i.e. 32 prequels out of the 406 observations), which is an indication of their lesser significance. Movies such as Star War 1, 2 and 3 may be considered as outliers rather than the norm.

The following set of robustness checks addresses the potential concern that our results may be sensitive to period-specific changes as would occur from one month to another or strategic holiday and/or summer month releases, all of which may influence the timing of movie releases. In column 4, we alter the main specification by including a set of month dummies. These modifications leave the original estimates essentially unchanged. We then included a dummy variable equaling one if the movie was released within a week's radius of major holidays (namely, Thanksgiving, Christmas, Memorial Day and July 4th), and zero otherwise (column 5). Next, we added a dummy variable that equals one for movies that are released in summer months and zero otherwise (column 6). We find no significant effect of either of these dummy variables.

The final robustness check addresses the possibility that series-specific effects play an important role in affecting time between sequels. Throughout this research, we searched for a proxy variable that might help control for these series specific effects, as using series fixed effects is not desirable since each series contains only one to five movies (this was the original motivation for the major studio variable). Such a variable was not found, however, as a robustness check, we ran series unique fixed effects and removed any control variable that did vary across different movies in the same series (e.g. *PreExisting.Universe*).

TABLE 3–Estimates from Log-Normal Regression: Time-to-Sequel Release in Days
(Robustness Checks)

Variable	Dependent Variable: log(Time)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	9.158*** (0.392)	8.861*** (0.416)	9.156*** (0.393)	9.078*** (0.439)	9.152*** (0.396)	9.120*** (0.396)	9.945*** (0.369)
log(Sales.of.Prev.Movie)	-0.148*** (0.043)	-0.102** (0.050)	-0.147*** (0.043)	-0.135*** (0.046)	-0.148*** (0.043)	-0.149*** (0.043)	-0.240** (0.096)
log (Budget	0.217*** (0.058)	0.220*** (0.057)	0.217*** (0.058)	0.208*** (0.058)	0.217*** (0.057)	0.216*** (0.057)	0.342*** (0.116)
Profitable.Prev.Movie	-0.591*** (0.128)		-0.593*** (0.127)	-0.557*** (0.134)	-0.591*** (0.128)	-0.500*** (0.130)	-0.454** (0.181)
L(Profit.of.Prev.Movie)		-0l.104*** (0.024)					
Different.Studio	0.524*** (0.090)	0.503*** (0.094)	0.525*** (0.091)	0.513*** (0.096)	0.523*** (0.091)	0.520*** (0.091)	0.316*** (0.120)
Major.Studio	0.366*** (0.085)	0.361*** (0.083)	0.369*** (0.086)	0.346*** (0.083)	0.366*** (0.085)	0.361*** (0.084)	-0.189 (0.242))
PreExisting.Universe	-0.346*** (0.081)	-0.360*** (0.082)	-0.348*** (0.081)	-0.300*** (0.085)	-0.345*** (0.080)	-0.346*** (0.080)	
Prequel			-0.047 (0.131)				
Holiday					0.009 (0.096)	-0.008 (0.099)	
Summer Release						0.052 (0.079)	

TABLE 3–Estimates from Log-Normal Regression: Time-to-Sequel Release in Days
(Robustness Checks)
(continued)

Variable	Dependent Variable: log(Time)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Year controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Genre controls	Yes	Yes	Yes	Yes	Yes	Yes	No
Rating controls	Yes	Yes	Yes	Yes	Yes	Yes	No
n-Stratification	Yes	Yes	Yes	Yes	Yes	Yes	No
Month controls	No	No	No	Yes	Yes	Yes	No
Series controls	No	No	No	No	No	No	Yes
Observations	406	406	406	406	406	406	406
Pseudo R^2	0.482	0.478	0.482	0.511	0.482	0.482	0.820

Notes: Robust standard errors are in parentheses below the coefficients. Statistical levels of significance are: *indicates significant at the 10 percent level; ** indicates significant at the 5 percent level; *** indicates significant at the 1 percent level. Pseudo R^2 calculated the same as in Table 2.

Once again, the results were largely unchanged, with a few coefficients increasing in magnitude and *Major.Studio* losing statistical significance. As the pseudo R^2 increases to 0.82, the likelihood of omitted variable bias changing direction or magnitude the results is low.

In summary, we find that, in almost all cases, the results obtained from the baseline specification persist across the various robustness checks as described above.

VI. Conclusion

The existing economic literature has neglected research into many aspects of movie sequels. This research focused on determining what variables influence the time between movie sequels and estimating the magnitude of their effect. Using survival analysis, time between sequels was modeled using an accelerated failure time model, and was determined that higher sales of the previous movie, a profitable previous movie, and a sequel based on a previous universe decreased time between sequels, whereas larger budgets along with being produced by a different studio or a major studio increased time between sequels. While the coefficient for both sales and budget were statistically significant, the extent of their economic significance was less noticeable.

References

Allison, Paul David. 1995. *Survival analysis using SAS: a practical guide.* Cary, NC: SAS Institute.

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories." *Sociological Methodology* 13: 61-98.

Basuroy, Suman, and Subimal Chatterjee. 2008. "Fast and frequent: Investigating box office revenues of motion picture sequels." *Journal of Business Research* 61 (7): 798–803. https://doi.org/10.1016/j.jbusres.2007.07.030.

Bohnenkamp, Björn, Ann-Kristin Knapp, Thorsten Hennig-Thurau, and Ricarda Schauerte. 2015. "When does it make sense to do it again? An empirical investigation of contingency factors of movie remakes." *Journal of Cultural Economics* 39 (1): 15-41. https://doi.org/10.1007/s10824-014-9221-6.

Cameron, Adrian Colin, and P. K. Trivedi. 2005. *Microeconometrics: methods and applications.* Cambridge, UK: Cambridge University Press.

Daniels, Bill, David Leedy, and Steven D. Sills. 1998. *Movie money: understanding Hollywood's (creative) accounting practices*. Los Angeles: Silman-James Press.

De Vany, Arthur S., and W. David Walls. 1997. "The Market for Motion Pictures: Rank, Revenue, and Survival." *Economic Inquiry* 35 (4): 783-797. https://doi.org/10.1111/j.1465-7295.1997.tb01964.x.

Marie Laure Delignette-Muller, and Christophe Dutang. 2015. "fitdistrplus: An R Package for Fitting Distributions. " *Journal of Statistical Software* 64 (4): 1-34. https://doi.org/10.18637/jss.v064.i04.

Einav, Liran. 2002. "Seasonality and Competition in Time: An Empirical Analysis of Release Date Decisions in the U.S. Motion Picture Industry." Stanford University Working Paper.

Fowler, Jennifer, Stuart J. Fowler, and Rush Hicks. 2013. "A Historical Investigation of Patterns in Sophomore Album Release." *MEIEA Journal* 13 (1): 61-74.

Grambsch, Patricia M., and Terry M. Therneau. 1994. "Proportional hazards tests and diagnostics based on weighted residuals." *Biometrika* 81 (3): 515-526. https://doi.org/10.1093/biomet/81.3.515.

Kreuzberg, Daniel. 2011. "Movies with Sequels." http://www.imdb.com/list/ls003495084/ (accessed March 28, 2015).

Nash Information Services, LLC. 1997. "The Numbers." www.the-numbers.com (accessed October 16, 2016).

R Core Team. 2016. *R: A language and environment for statistical computing.* Vienna, Austria.

Rodriguez, Germán. 2010. "Survival Analysis." Princeton University, Summer. http://data.princeton.edu/pop509/ParametricSurvival.pdf (accessed October 16, 2016).

Rodriguez, Germán. 2007. "Lecture Notes on Generalized Linear Models: Chapter 7." Princeton University. http://data.princeton.edu/wws509/notes/ (accessed October 16, 2016).

Therneau, Terry M. 2015. "A Package for Survival Analysis in S." https://CRAN.R-project.org/package=survival (accessed October 16, 2016).

Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling survival data: extending the Cox model.* New York: Springer

Walls, W. D. 2009. "Screen wars, star wars, and sequels: Nonparametric reanalysis of movie profitability." *Empirical Economics* 37 (2): 447-461. https://doi.org/10.1007/s00181-008-0240-z.

Appendix TABLE 1–Variable Definitions

Variables	Definition	Source
Time	The number of days between the release of the movie in question and the previous movie in the series.	Numbers
Sales.of.Prev. Movie	U.S. Box Office sales of the previous movie in the series in millions of real dollars. Base period 1980-1982	Numbers
Budget	Movie budget in millions of real dollars. Base period 1980-1982.	Numbers & IMDb
Profit.of.Prev. Movie	U.S. Box Office sales minus movie budget in millions of real dollars. Base period 1980-1982	Numbers & IMDb
Profitable. Prev.Movie	Whether *Profit.of.Prev.Movie* is positive: 1 if it was profitable, 0 if it was not.	Numbers & IMDb
PreExisting Universe	Whether the concept of the movie is based upon a universe from a pre-existing work such as a book, comic book, historical events, etc. Does NOT include previous movies as these are sequels. 1 if yes, 0 if no.	Numbers
Different. Studio	Whether the sequel is being produced by a different studio as the last movie in the series: 1 if it is a different studio, 0 if it is the same.	Numbers
Sequel Number	The number of the sequel in the series	IMDb
Major.Studio	Whether the movie was produced by one of the six major studios: 1 if yes, 0 if no.	Numbers
Rating	The rating of the movie. The ratings include G, PG, PG13, R, and NC17. This variable is represented as a vector of dummy variables to control for the fixed effects of ratings with PG being the one excluded.	Numbers
Year.of.Prev. Movie	The year the previous movie in the series was released domestically. Represented as a vector of dummy variables to control for fixed effects with 2000 being the year excluded.	Numbers
Month	The month the movie was released.	Numbers
Prequel	Whether the movie was a prequel: 1 if yes, 0 if no.	Numbers
Genre	The genre of the movie. This variable is represented as a vector of dummy variables to control for fixed effects with Action being the genre excluded. The genres include Horror, Action, Adventure, Comedy, Family, Animation, Superhero, SciFi, Fantasy, Drama, Mystery, Romance, Music, Sport, Crime, Biography, and History	IMDb

Notes: Numbers refers to The Numbers by Nash Information Services, LLC. IMDb refers to the Internet Movie Database. Dollars were adjusted for inflation using BLS Consumer Price Index with 1980-1982 as the base period.

Appendix TABLE 2–Correlation Matrix

Variables	1	2	3	4	5	6	7
1 Sales.of.Prev.Movie	1.000						
2 Budget	0.635***	1.000					
3 Profitable.Prev.Movie	0.366***	0.024	1.000				
4 Profit.of.Prev.Movie	0.667***	0.216***	0.762***	1.000			
5 Different.Studio	-0.273***	-0.174***	-0.121**	-0.222***	1.000		
6 Major.Studio	0.302***	0.333***	0.054	0.125**	-0.174**	1.000	
7 PreExisting.Universe	0.175***	0.334***	-0.067	-0.028	-0.041	0.098***	1.000

Notes: Statistical levels of significance are: * indicates significant at the 10 percent level, ** indicates significant at the 5 percent level, *** indicates significant at the 1 percent level.