International Journal of Operations Research and Information Systems
Volume 11 • Issue 4 • October-December 2020

A Parsimonious Predictive Model of Movie Performance:
A Managerial Tool for Supply Chain Members

Mustafa Canbolat, SUNY Brockport, Brockport, USA
Kyongsei Sohn, SUNY Brockport, Brockport, USA
John T. Gardner, SUNY Brockport, Brockport, USA

ABSTRACT

In this paper, the authors develop a parsimonious model that offers early prediction of potential success of a movie. In order to achieve this, a broad look at the drivers of movie success is required. Supply chain members will be making decisions regarding movie popularity with regard to licensing contracts, forecasting toy purchases, cross-promotions, etc. at varying times before a movie is released. A simple forecasting approach using publicly available data could improve supply chain decision making. Prior literature suggested that the virtual movie stock market, HSX, was a good predictor. Using a small set of variables including view counts, likes, and dislikes did offer some predictive value. However, HSX produces a forecast that dominates prior models while using a single readily available public data. Further, the HSX-based prediction showed consistency and convergence across a significant breadth of time.

KEYWORDS

Forecasting, HSX, Movie Revenue Prediction, Supply Chain Management, Virtual Stock Markets

INTRODUCTION

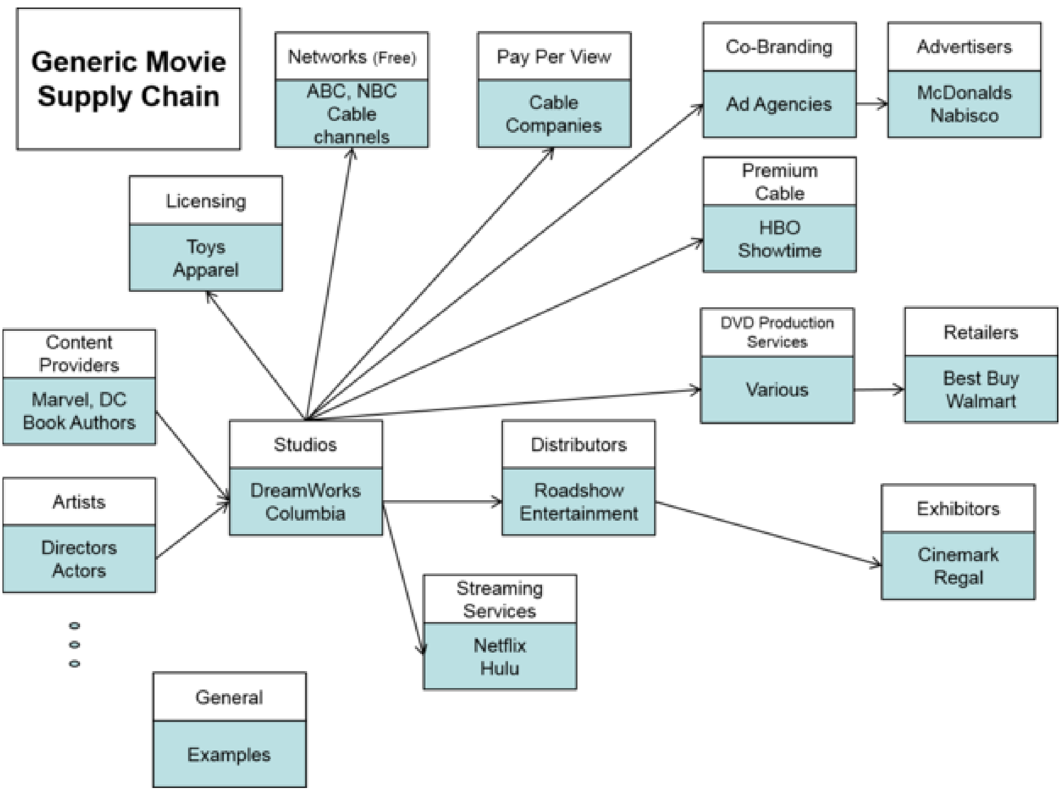
The success of movies contributes to the success of many businesses such as toy manufacturers utilizing movie tie-ins, mass merchandizers selling DVDs, consumer product companies utilizing co-branding and advertising tie-ins, etc. Many supply chain decisions would benefit from some additional data-based knowledge regarding movie success. These decisions occur over a broad time horizon generally before the release of the movie. The academic literature offers multiple approaches to modeling movie success as detailed below. The models are typically complex, using data difficult to obtain.

There appears to be a need for a simple predictive model accessible to purchasing managers, advertising executives, brand managers and other supply chain managers not versed in forecasting complexities. The ideal models will have few variables, data that can be conveniently obtained at low or minimal cost, minimal complexity, and robustness with respect to lead time before release. For example web based data that can be accessed easily and immediately would be ideal. Simple is better.

DOI: 10.4018/IJORIS.2020100103

Supply chain members vary considerably. Primary members might include movie production companies, distributors and exhibitors. Figure 1 depicts the range of potential channel members both primary and peripheral. The model developed here would not likely be of interest to primary members, as they would do more extensive, more complex, more frequent analysis of a movie's potential success. Peripheral members would often only need a single early predictor. For instance a purchasing agent for mass merchandiser has to choose a set of action figures for the toy section. This decision is likely made relatively quickly as there are many SKU (Stock Keeping Units) slots to fill. Additionally, any given decision may reflect only a small economic stake as a result of an individual movie-related decision. An example might be a choice of kid's meal toy by a fast food chain. That choice is important but is in no way a make-or-break decision.

Figure 1. Generic movie supply chain



The modeling objective requires identifying a set of variables that consistently predict future movie success which will be measured by US box office revenue. An individual supply chain member can translate the box office revenue into an estimate of their chances for success based on their own history of prior decision making and related box office results. Many of these decisions are made repeatedly, such as cross-promotion choices made by cereal manufacturers.

Many variables have been used in previous studies, however they vary significantly with respect to their availability and the ease of access for practical use. Many of the predictors of movie revenue used in earlier studies included variables that are collected just prior to or after a movie's release. In addition to variables used in earlier studies, a search for new variables was undertaken.

A search was conducted for new and existing variables which might provide foundation for such a forecasting model. A supply chain manager could look up movie trailer views, trailer likes and dislikes. These three variables are the new variables not found in the literature. An existing variable, HSX virtual stock price, appeared to capture many aspects related to movie success reflected in prior models. HSX is a virtual stock market game in which members trade movie stocks with the price based on movie revenues denominated in millions of dollars of US domestic box office revenue. More details on HSX will be given below. These four variables can be obtained from two websites at any time therefore they are extremely accessible.

There are multiple uses for this prospective model. After a licensing deal is made a toy manufacturer may need to forecast for the first manufacturing run of a movie-based toy. A check on any jointly developed forecast of success built collectively by the supply chain may need a reality check. An integrated supply chain dominated by the major producer could result in biased forecasts. There may be more cases where an independent, simple, fast, reliable forecasting technique would aid decision making.

The discussion is organized as follows. Section two gives a background and literature review. Section three details the model development process. Section four reviews the results. Section five offers conclusions.

BACKGROUND

To build an effective and efficient forecasting model for the smaller players in the movie supply chain there are a number of considerations. The economic importance of the movie supply chain should be reviewed. Prior movie performance predictive models should be considered. In order to identify new variables, earlier input variables need to be identified and evaluated for their appropriateness for the current objective.

Movie Supply Chains

Movie supply chains are important in just the licensing aspect alone. See Figure 2. Not all of these licensing deals involve two large, central supply chain members. For the peripheral members, the deals may be smaller dollar values but still very significant.

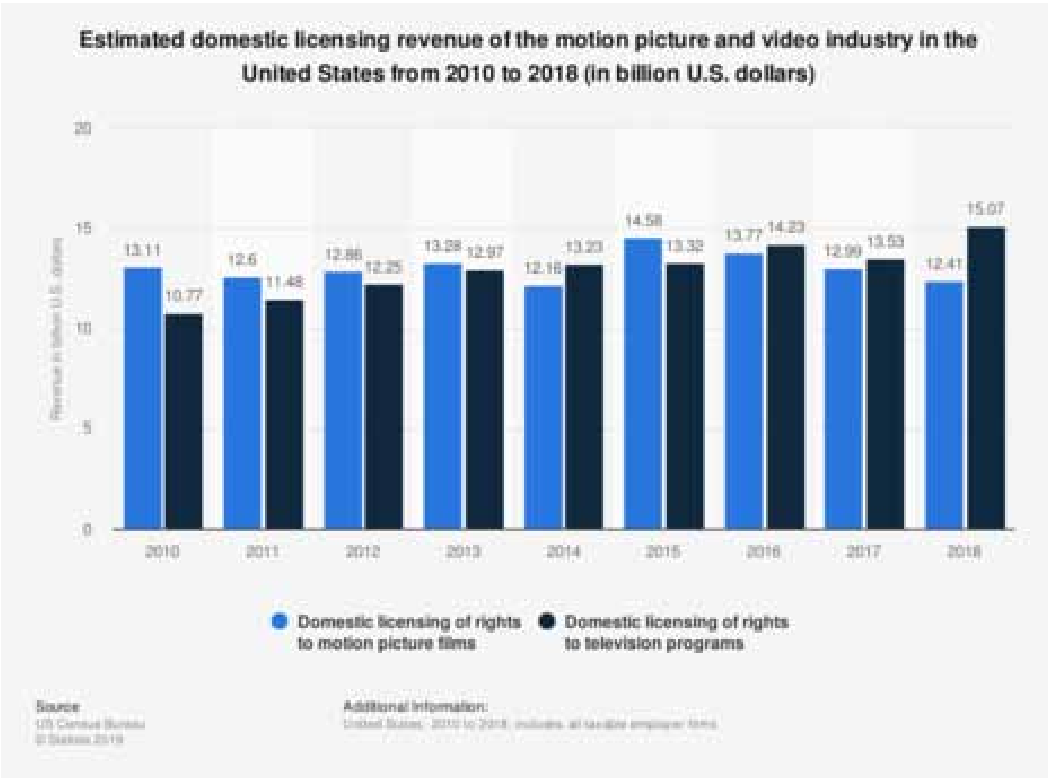
Most supply chain research on forecasting within the chain is focused on collaborative efforts between members with significant volume business. Examples of such research include Aviv, Y. (2001), McCarthy and Golicic (2002) and Helms et al. (2000). Supply chain forecasting research has also focused significantly on the Bullwhip effect. See for instance Chen et al. (1999, 2000). In each case, the collaborative aspects are emphasized. Most business-to-business interactions do not, however, involve significant collaboration, particularly with respect to peripheral members (Cooper et al. 1997). There appears to be value in taking a supply chain view of forecasting in the movie industry. Movie forecasting has a significant literature that does not specifically address supply chain issues.

Movie Revenue Forecasting Literature

Movie revenue forecasting has been examined by a number of researchers. The models use many approaches and a wide range of variables. The approaches and the variable choices are summarized in Table 1 and measures of fit and the timing of the availability of the variables are covered in Table 2. For our objectives, the variables should be available early.

The approaches include regression analysis, a dynamic artificial neural network, a machine learning analysis and others. These models vary significantly in their complexity and potential managerial acceptance. The newer methods and more complex methods would likely have to overcome some managerial scepticism. Not all of the studies were focused on predictive accuracy. Pennock et al (2001) focused only on evaluating the put-call parity of HSX and did not use other variables.

Figure 2. US Domestic Licensing; 2010 to 2018; includes all taxable employer firms (US Census Bureau, 2019)



Similarly, Elberse (2007) evaluated the impact of star power on movie revenues using an event study where events are considered such as casting announcements.

The two main issues in Table 1 are the quality of fit and how early the model can be run. Model fit measures cannot be directly compared across model types. Looking at R^2 values indicates the models vary significantly in terms of percent of variance explained. Note that Simonoff and Sparrow (2000) achieved an R^2 of 0.968 by including the first week's gross revenue. Typically, movie-related supply chain agreements are completed well before the end of the first week of the movie's release.

Table 2 offers a summary of the appropriateness of the prior models for the purpose of the proposed forecasting model. Each model has some challenges regarding our search for a simple and accessible approach available to peripheral supply chain members. Models have varying degrees of problems with simplicity, data availability, timing, or potential managerial acceptance.

The above research synthesized in Table 3 have furthered the understanding of what drives eventual movie revenues. By using a multiplicity of approaches and various sets of variables the models presented in the research above gives a more complete picture. None of these cited articles claimed to offer smaller supply chain members a simple and accessible forecasting tool that could be relied on across a spectrum of time. They lack documented robustness with respect to lead time before a movie's release.

Managerial Timeline

Various supply chain members will have different timelines for their movie-related decisions. An ideal model will work for a variety of lead times before a movie's release. A product placement decision must be made very early while a small, nimble toy manufacturer may be able to wait significantly

Table 1. Movie revenue forecasting/prediction models

Author(s)	Model	Dependent Variable	Included Variables							
			Movie Rating	Genre	Screen Count	Star Value	Seasonality	Sequel	Production Budget	Other
Ghiassi et al (2015)	Dynamic artificial neutral network	Box-office revenue	✓	✓	✓	✓	✓	✓	✓	Competition, runtime, pre-release advertising expenditures, special effects, etc.
Hur et al (2016)	Machine learning algorithms with ISM	Box-office revenue	✓			✓	✓			Machine learning with viewer sentiment score, etc.
Simonoff & Sparrow (2000)	Regression	Ultimate Gross revenue	✓	✓	✓	✓	✓	✓	✓	First week gross revenue, motion picture award nominations & wins
Marshall et al (2013)	New product diffusion models	Movie attendance	✓	✓	✓	✓	✓	✓	Advertising spending	Local newspaper, competition, reviews, etc.
Foutz and Jank (2010)	Functional Shape analysis model	Opening weekend revenue	✓	✓	✓	✓	✓	✓		HSX Friday closing price
Huang et al (2015)	Log linear	Cont. comp. ROI (profit of low-budget vs. big-budget films)	✓		✓	✓	✓	✓	DVD ad spending, pre-release ad spending	Positive critics review, positive consumer review, major/ mini-major studio
Spann & Skiera (2003)	Regression/ Forecasting accuracy	Movie revenue		✓	✓					HSX Friday closing price, Box Office Mojo predictions
Karniouchina (2011a, 2011b)	Log-log	Movie revenue	✓	✓	✓	✓	✓	✓	✓	Word of mouth, prediction, HSX closing price
McKenzie (2013)	Log-log	Movie revenue	✓	✓	✓	✓	✓	✓	✓	HSX prices before opening weekend
Pennock et al (2001)	Box office forecast accuracy	Four-week movie revenue								HSX Friday prices, Box Office Mojo's revenue prediction
Elberse (2007)	Constant-Mean-Return model	Movie theatrical revenue				✓				HSX Friday closing prices after a cast announcement, StarBond market information

Table 2. Model fit and timing

Author(S)	Model	Dependent Variable	How Early?	Model Fit/R^2	Relevant Key Findings	Country
Ghiassi et al (2015)	Dynamic artificial neutral network	Box-office revenue	Pre-production model	N/A	DAN2 model improves forecasting movie revenue	USA
Hur et al (2016)	Machine learning algorithms	Box-office revenue	Not specified (N/A)	N/A	viewer sentiment score	South Korea
Simonoff & Sparrow (2000)	Regression	Log gross revenue after the first week	Post release	.446 ~ .968	First week gross revenue	Mainly focus on USA
Marshall et al (2013)	New product diffusion models	Movie attendance	3 weeks of data	0.161 ~0.280	Bass, Sawheny & Eliashberg model parameters	Chile
Foutz and Jank (2010)	Functional Shape analysis model	Opening weekend revenue	Most recent 54 trading weeks data	MAPE (5.28 – 19)	HSX (Virtual Stock Market price)	USA
Huang et al (2015)	Log linear	Log ROI	Post release	0.554-0.820	DVD ad spending, pre-release ad spending, Ratings	USA
Spann & Skiera (2003)	Regression/ Forecasting accuracy	Movie revenue	Friday before opening	.859-.861 MAPE (36.48 – 40.62)	HSX prediction, Box Office Mojo predictions	USA
Karniouchina (2011)	Log-log regression	Movie revenue	Friday before opening	.86-.87	HSX predicted Market price, movie/star buzz	USA
McKenzie (2013)	Log-log regression	Opening box-office revenue	Friday before opening	.893 -.895 MAPE (.273-.299)	HSX predicted Market price	USA
Pennock et al (2001)	Box office forecast accuracy	Four-week movie revenue		3.57 MSE, 31.5% MPE	Friday morning's price of HSX	USA
Elberse (2007)	Constant-Mean-Return model	Movie theatrical revenue		N/A Event study	HSX stock prices, StarBond market info	USA

longer for a licensing agreement. An ideal model will include assurance that the process can be used both early as well as offer flexible timing of use.

MODEL DEVELOPMENT

In order to meet our goals for the research, three hypotheses will be tested. Two different datasets are employed. Beyond the three testable hypotheses there are two additional research questions which cannot be specifically tested through statistics. These questions will follow the hypotheses tests below.

Table 3. Managerial accessibility

Author(S)	Model	Note on Appropriateness of Current Objective: Simple and Publically Available Forecasting Model
Ghiassi et al (2015)	Dynamic artificial neutral network	Not simple, not early, not managerially intuitive
Hur et al (2016)	Machine learning algorithms with ISM	Not simple (review sentiments analysis, non-linear machine learning algorithms, classification and regression tree, support vector regression, etc.), not early, not managerially intuitive
Simonoff & Sparrow (2000)	Regression	For better prediction the model includes post release revenues, clearly not early
Marshall et al (2013)	New product diffusion models	Available after movie release (Bass/Sawheny & Eliashberg model parameters), clearly not early
Foutz and Jank (2010)	Functional Shape analysis model	Not a predictive model, not simple, not early, not managerially intuitive
Huang et al (2015)	Log linear	Available after movie release, clearly not early
Spann & Skiera (2003)	Forecasting accuracy	Simple, publically available, early, free, but lacks managerial intuitions, flexible timing for decision making not addressed
Karniouchina (2011)	Log-log	Available after movie release, clearly not early
McKenzie (2013)	Log-log	Available after opening weekend, clearly not early
Pennock et al (2001)	Box office forecast accuracy	Not a predictive model
Elberse (2007)	Constant-Mean-Return model	Not a predictive model (event study)

New Variables

One of the goals of the research is to find new variables that might work well for our supply chain forecasting approach. Trailers communicate information related to multiple variables as depicted in Table 1. For instance, the movie genre is obvious, actual star value should be clearly visible, and the relationship with prior movies in a series (sequel) can be evaluated by the viewer. A variable that captures multiple prior proven influencers on movie revenue could help simplify a forecasting model. Internet use of trailers is clearly a popular venue given that trailers get up to millions of views. While trailers are also viewed in theaters or television spots only Internet views meet our objectives of simplicity and accessibility for supply chain members. In a similar vein, likes and dislikes capture the social media aspect of movie revenue forecasting. Hur et al. (2016) demonstrated that forecasting accuracy can be enhanced using review sentiments which often include positive, neutral, and negative opinions. Ding et al. (2017) used Facebook likes to understand the impact of social media on box office. The authors revisited the Social Impact Theory (Latane and Nida, 1980) and considered Facebook "likes" as the first of the three social forces: number, immediacy and strength. We believe YouTube "views", "likes", and "dislikes" are similar to social media sentiments and Facebook "likes" and they can be considered as easy to reach additional social metrics which may help supply chain members make better decisions.

In relation to using these new variables the following hypothesis is proposed:

Hypothesis 1: Views, likes and dislikes can predict movie success.

To test our hypothesis, we wrote a small script in Google Sheets which made automated visits to a popular YouTube channel named Movieclips Trailers that delivers the latest trailers for upcoming movies. The channel has more than ten million subscribers. We observed that the channel posts the trailers on the same day of official trailer release. The script collected the number of views, likes, and dislikes on 120 movie trailers posted on this channel every day for about a year. We observe the shapes of movie trailer cumulative view growth across many movies and determined that a two-week window gave sufficient information for our purposes. This was done by visual observation of growth curves. As some movies have multiple trailers, we chose to include only the first official trailer for each movie. The sample collection started on Oct 27, 2016 and ended around the same time the following year. A small number of trailers may have been missed due to server related difficulties during the data scraping process. The resulting sample size was 71.

Revenue data is collected primarily from Box Office Mojo <www.boxofficemojo.com> and when necessary supplemented by The Numbers <www.the-numbers.com>. Both of these websites track domestic box office receipts. The analysis was limited to domestic revenues.

As our objective is early prediction of movie success, we looked at the cumulative views, likes, and dislikes of these movies two weeks after their first trailers release date and see if these numbers have any relationship with the movie revenues within two weeks or six weeks of movie release. The descriptive statistics of the variables are in Table 4 and the correlation matrix is in Table 5. The descriptive statistics indicate that all the variables we considered have right skewed distributions indicating the presence of a small number of movies that drive the mean value of views, likes, dislikes, and revenues up.

Table 4. Descriptive statistics

	Two-Week Views	Two-Week Likes	Two-Week Dislikes	Two-Week Revenue	Six-Week Revenue
Mean	1,405,068	18,641	1,003	$52,951,738	$73,086,883
St. Dev	2,054,786	24,725	2,026	$74,997,096	$102,323,21
Median	784,399	8,163	478	$28,537,985	$37,623,226
Min	15,452	47	7	$49,982	$96,722
Max	13,477,412	118,698	14,984	$375,378,70	$520,200,08

Table 5. Correlation Matrix

	Two-Week Views	Two-Week Likes	Two-Week Dislikes	Two-Week Revenue	Six-Week Revenue
Two-week views	1	0.963102	0.803564	0.379148	0.415052
Two-week likes		1	0.752073	0.471373	0.51244
Two-week dislikes			1	0.316911	0.319985
Two-week revenue				1	0.962281

The correlation matrix indicates a presence of multi-collinearity as a result of a strong positive relationship among the views, likes, and dislikes. This issue will not affect the predictive power of a model that uses a mix of these variables but we will not be able to draw any conclusion from the objective coefficients. Further, the strong correlation between the two-week revenue and six-week revenue confirms that the earnings within two-weeks will be indicative about the movie's eventual success.

We used a number of different ordinary least squares models presented in Table 6 to test our hypothesis. All variables excluding the binary ones are log scaled to address the issue of skewness. The HSX value is scaled in millions of dollars and the revenue values are similarly scaled. The following notation is used:

- *REVij*: Cumulative movie revenue *j* weeks after movie *i*'s release date.
- *VIEWSit*: Cumulative number of trailer views *t* weeks after the trailer release date of movie *i*.
- *LIKESit*: Cumulative number of likes *t* weeks after the trailer release date of movie *i*.
- *DLIKESit*: Cumulative number of dislikes *t* weeks after the trailer release date of movie *i*.
- *Ri*: Dummy variable equal to one if the MPAA (The Motion Picture Association of America) rating of movie *i* is R, zero otherwise.
- *ACTi*: Dummy variable equal to one if the genre of movie *i* includes action, zero otherwise.

Table 6. Ordinary least square regression models

Model 1	$LOGREV_{i6} = \beta 0 + \beta 1 LOGV\,IEWS_{i2} + \varepsilon i$
Model 2	$LOGREV_{i6} = \beta 0 + \beta 1 LOGLIKES_{i2} + \varepsilon i$
Model 3	$LOGREV_{i6} = \beta 0 + \beta 1 LOGDLIKES_{i2} + \varepsilon i$
Model 4	$LOGREV_{i6} = \beta 0 + \beta 1 LOGV\,IEWS_{i2} + \beta 2 LOGLIKES_{i2} + \beta 3 LOGDLIKES_{i2} + \varepsilon i$
Model 5	$LOGREV_{i6} = \beta 0 + \beta 1 LOGV\,IEWS_{i2} + \beta 2 LOGLIKES_{i2} + \beta 3 Ri + \beta 4 ACTi + \varepsilon i$

The results containing the coefficients and the significance of the variables used in each model along with the coefficient of determination values are provided in Table 7. The first three models look at the relationship between the cumulative movie revenue over a six weeks period and trailer

Table 7. Coefficients and significance results

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	7.953**	10.441***	13.397***	21.257***	-3.515
$LOGVIEWS_{i2}$	0.664***			-1.841**	-1.906**
$LOGLIKES_{i2}$		0.712***		2.188***	1.991***
$LOGDLIKES_{i2}$			0.568**	0.106	
Ri					-1.229**
ACTi					1.437**
(Adj.) R-Square	0.171	0.263	0.104	0.319	0.426

. Significance at the 10% level,
* Significance at the 5% level,
** Significance at the 1% level,
*** Significance at the 0.1% level.

views, trailer likes, and trailer dislikes two weeks after the trailer is released. The results show that although the variables are significant thus they support Hypothesis 1, the R-square values show a low level of explained variability which implies a lack of predictive ability. The most significant variable is $LOGLIKES_{i2}$, which is followed by $LOGVIEWS_{i2}$.

Model 4 uses all three variables at the same time but because of the high level of multi-collinearity present in the model, the coefficients became unstable. Nonetheless, the adjusted R-square value is improved. In Model 5 we eliminated $LOGDLIKES_{i2}$ from the model as it was not significant and incorporated all the other variables including the MPAA rating and genre which are the control variables. The model reaches the highest adjusted R^2 value of 42.6% but loses some degrees of freedom. The sample size is still large enough to have an accurate predictive model as it was suggested in Harrell (2001) that minimum of 10 samples per variable is needed in an ordinary least square regression model. The results show that early prediction of movie revenues using movie trailer data is possible but the power of prediction is not as desired.

Given that our findings for these variables do not look as promising to predict eventual movie revenue, additional investigation is required. By adding the values derived from the HSX virtual stock market game to the above variables how much can be gained?

Adding HSX

Three of the prior studies referenced in Table 1 included HSX however none of them used HSX exclusively or with the above variables. Elberse (2007) and Elberse and Anand (2007) examined the effect of events and advertising on HSX, but did not have the current focus on supply chain forecasting. Our goals include ease of accessibility and early accessibility. A model that includes the above variables and HSX should be tested. Hypothesis 2 follows:

Hypothesis 2: HSX along with views, likes and dislikes can predict movie success.

Our data collection using Google Sheets that we described earlier included automated visits to the HSX website and we collected the daily stock prices of 120 movies over a period of one year. But because of the elimination of follow up trailers from our dataset that we discussed before, the combined dataset including the YouTube data and HSX data ended up having the same sample size of 71. We used the HSX stock values of movies two weeks after the first trailer release (*HSXi*2) similar to the process we have done for the views, likes, and dislikes. The new model incorporating the HSX variable is as follows:

Model 6: $REV_{i6} = \beta 0 + \beta 1 VIEWS_{i2} + \beta 2 LIKES_{i2} + \beta 3 HSX_{i2} + \beta 4 Ri + \beta 5 ACTi + \varepsilon i$

The regression results show that Hypothesis 2 is strongly supported with an 84.5% adjusted R-square value. However, the HSX is the only variable that is significant and is dominating the predictive ability of the model.

HSX Alone

The above empirical result may be an indication that HSX is absorbing the information explained by all the other variables and coming out as a sole predictor which has more predictive power than all the other social media related variables that we used. Hypothesis 3 captures this idea.

Hypothesis 3: HSX alone can predict movie success.

The hypothesis is tested with the following model:

Model 7: $REVi6 = \beta 0 + \beta 1 HSXi2 + \varepsilon i$

Even though the model eliminated four explanatory variables, it did not lose any prediction power. The model has an R-square of 83.6% and returns a significant estimate for HSX at the 0.1% level. The estimated regression equation is:

$$REVi6 = -10.054 + 1.068 HSX_{i2}$$

To further validate the use of HSX as a sole predictor, we collected daily HSX values of 150 recent movies from the HSX website for over a year. Similar to the process we have employed for the earlier dataset, we matched the six-week revenue data from boxofficemojo.com with these HSX values. For the sake of completeness, for movies that did not run the full six-week period we used the delist price of HSX which represents four-week revenue.

The results from the larger dataset confirmed the significance of HSX but with a slightly lower R-square value of 78.1% when using the HSX value two weeks after the trailers release dates to predict six-week revenue.

In conclusion, HSX alone appears to be a better predictor when accounting for simplicity and accessibility as our goal is a simple and accessible predictor therefore testing the maximally simple approach was attractive. HSX has a large advantage in that the value at closing is the revenue stated as one dollar for each million dollars of revenue and it should be instructive to test HSX predictive value alone. The HSX stock values change in real-time when players are competing against each other to better predict that value at any point in time prior to release. This leaves two important questions for using HSX as a predictor.

Model Choice

Which of the above models works best? The percent variance explained thoroughly dominated by HSX as a predictor. The supply chain management related goals at the outset of the research would all point to choosing a simple regression model using only HSX. HSX value is available early and is publicly available. In addition, it can be used directly as an estimate of future movie revenues denominated in millions.

How early and how well does the final model work? Supply chain partners will want to know how good a prediction would be if they simply choose to look at the HSX value at the point of time when they need to make a decision concerning an upcoming movies success. In order to help them have confidence in HSX as a single predictor further analysis is required.

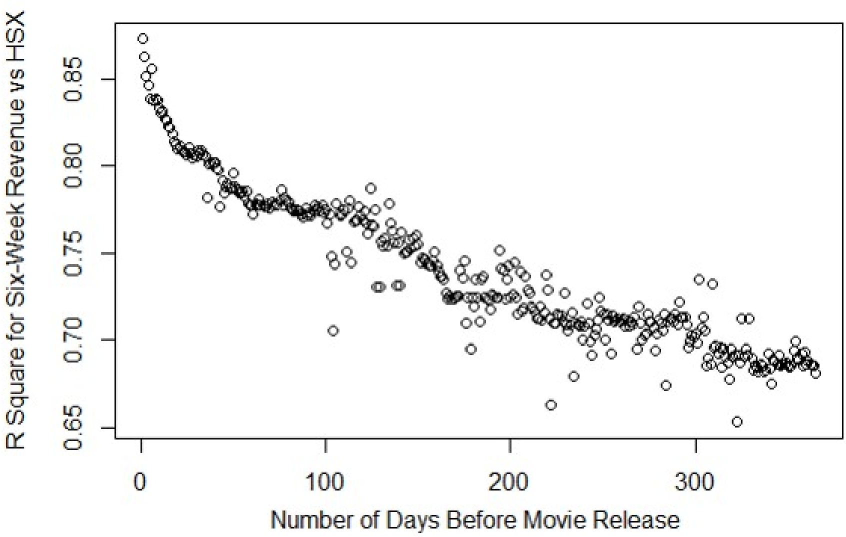
Predictive Dynamics

We tested Hypothesis 3 using a timeframe matching our analysis of Hypothesis 1 and Hypothesis 2. This is a single point in time, two weeks after the initial trailer is released, which is typically months ahead of the release date. However, supply chain members may need to use this information at various points of time prior to a movies release. By looking at the quality of prediction at any given day prior to a movies release, a comparative measure of quality of fit can be observed. Therefore, an examination of the degree of improvement in predictive power as time progresses is worthwhile.

To this end for 365 days, for each day 150 movie HSX values were regressed against their six weeks total revenue value. Thus 365 regressions were used as input to a graph of R-square values showing how they converge as time progresses. Multiple timings of revenues were examined including two, four, and six weeks as a dependent variable. All yielded very similar results as all three strongly correlate with each other. Six week revenue is preferred as it would capture a good measure of both wide release and limited release movies. See Figure 3.

Figure 3 indicates that the predictive value is quite strong from a hundred days prior to release. The R-square values are indicating a convergence between the prediction and the actual revenue. This is a purely subjective observation. No statistical method was found that tests the rate of convergence of the R-square values. The R-squares subjectively at about 100 days before release compare very favorably to most measures of fit described in Table 2. The estimates prior to 100 days are likely to represent a very good and early estimate that a supply chain manager could obtain with reasonable effort and resources. We used the release date as the end of our analysis since we believe supply chain members typically make movie-related decisions well prior to release.

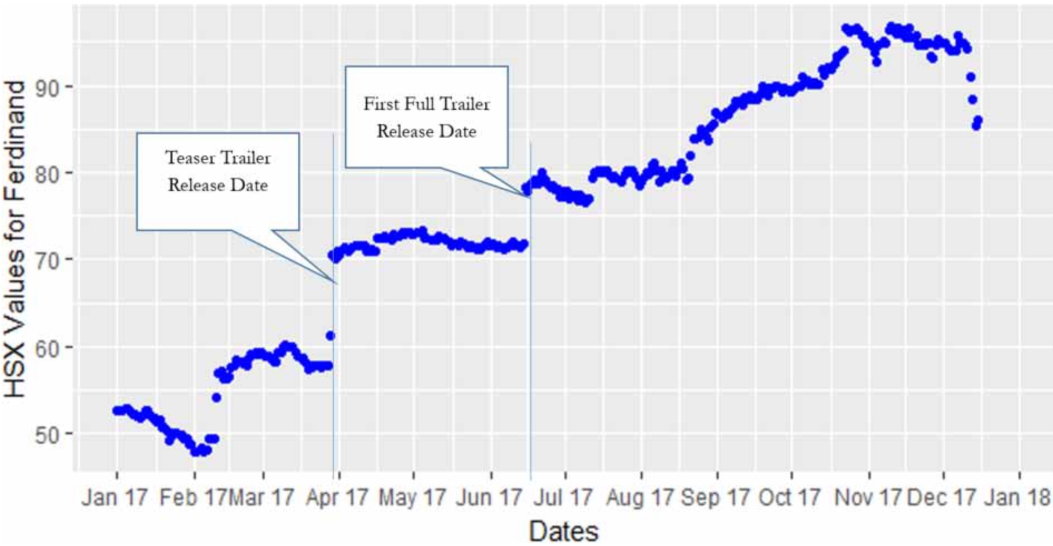
Figure 3. The change in R-square over time



An animated movie, Ferdinand, is chosen to illustrate supply chain managers use of HSX as a predictor of a movie's success. The time series of HSX values for this movie is illustrated in Figure 4. Ferdinand's first trailer is released on March 28, 2017 as a teaser trailer and the first official trailer is released on June 14, 2017. Note the jumps on these two days in the figure. These visually observed structural changes may be the HSX players' response to the trailers and result in the absorption of trailer related information to HSX values. The movie is released on Dec 15, 2017 and the two-week revenue was $50.79 (million). The movie was delisted at the price of $70.45 (4-week revenue) and six week revenue reached at $80.61 (million). The final domestic box office revenue is $84.41 (million). A supply chain manager would have a reasonable estimate of the movie's success at multiple points in time and as time progress, the estimate generally improved.

Two weeks after the first full trailer is released, the HSX value for Ferdinand was 78. That means from the perspective of a toy manufacturer, at two weeks after its first trailers release, they would project a final box office revenue of 78 million dollars if they use HSX as a direct measure to predict the movies success. When the managerial cost of additional complexity overwhelms the economic value of slightly better prediction, they should just use HSX. However, a statistically better estimation may be possible using Model 7. The actual use of Model 7 would require sophisticated statistical analysis specific to the time frame of their decision.

Figure 4. The change in HSX Values for Ferdinand over time



Understanding this balance between the complexities inherent in our statistical analysis and the cost of risks over time for using a direct simple approach offers a rich opportunity for future research.

CONCLUSION

This section is organized as follows: overall conclusions, limitations, and future research.

Overall Conclusion

The proposed new variables are of limited value and are dominated by HSX. Note the quick incorporation of trailer related information; trailer quality, trailer popularity and other embedded information into the HSX values as suggested by the discontinuities in the above graph. Karniouchina (2011a) has made the case that virtual markets are efficient in incorporating demand information. The operational complexity of using trailer-related variables does not appear to be justified.

We found HSX alone offers extremely parsimonious and effective predictor of movie revenue potential. Specifically, how it is used is an open question. Because HSX is designed in such a way that stock value at any point in time represents the players collective beliefs concerning the final revenue denominated in millions of dollars, the stock value may be used directly. Alternatively, the HSX value can be transformed into a somewhat better estimate through the use of Model 7 above. The added value that the model use represents is questionable. Additional research is indicated.

Up to three months prior to release the predictive power of HSX alone model is competitive with the much more complicated models in Table 2. In addition, all of the prior models showed some problems for supply chain use. Our goals for simplicity, parsimony and early availability of publicly available data are best met through the use of the simplest of models.

Limitations

This study is for domestic supply chain decision-making. International or specific country decisions will have to use HSX with a great deal of caution. Worldwide revenue estimation using HSX offers a potential for significant errors. There are many causes for this such as: different cultural acceptance rates, the quality of translation or dubbing, governmental acceptance of movie distribution, etc. Cross-cultural acceptance of movies varies significantly across country boundaries (Craig et al. 2005).

In addition, there are data-based limitations including sample size, the settling of the revenues over time, digital shift in viewing habits, omitted variables, revenue accuracy (occasionally multiple sources for movie revenue vary), among others.

Many supply chain decisions will not justify major investments in complex data collection and mathematical modeling. However, primary members of the supply chain making a decision or peripheral members making a decision with potential for disastrous results should likely use more complex and expensive approach to forecasting. Supply chain managers should be warned that any competitive effect will not be incorporated into HSX values until the movie release schedules for major studios are finalized.

The movie industry is in a period of transition. Streaming-only movies may become more significant. Streaming success may become a bigger part of a supply chain managers' decision. For instance, a toy manufacturer thinking about an action figure must consider the effect of streaming viewing on demand that is not included in the HSX values. Children often want to watch a movie over and over which may change the demand for movie-related toys.

Future Research

Figure 3 depicts an increasing percent of variance explained over time. The relationship between the aggregate improvement of the R-square and the individual movie predictions based on individual HSX values can be further examined. Understanding more about the dynamics and the reliability of this relationship may provide additional insights. A means of comparison across predictors would enhance supply chain confidence in their decision making. The influence of externalities may lead to discontinuities as illustrated in Figure 4. Event studies or mathematical modeling of dynamics of discontinuities could offer a valuable contribution for the literature. In the current case, a trailer release or a key piece of entertainment news appears to be incorporated quickly into the HSX value based on the example. With a larger sample, an analysis of discontinuities may confirm their importance, clarify their influence, and potentially predict their consequences.

Final Conclusion

The simple linear model was supported and could be employed with consideration for the derived coefficients. The prediction of a movie's revenue involves adding 6.8% to the value of HSX and subtracting $10.05 to the estimate of earnings in millions if a manager decides to use the HSX value at two weeks after the first trailer's release date. If this manager needs to make a significant but not critical decision related to a movie's success at any given time, the HSX value itself likely offers a good estimate without adjustment. Adding complexity to the prediction process may not be justified.

REFERENCES

Alster. (n.d.). It's just a game, but Hollywood is paying attention. *The New York Times*.

Aviv. (2001). The effect of collaborative forecasting on supply chain performance. *Management Science, 47*(10), 1326–1343.

Chen, Drezner, Ryan, & Simchi-Levi. (1999). *The bullwhip effect: Managerial insights on the impact of forecasting and information on variability in a supply chain.* Academic Press.

Chen, F., Drezner, Z., Ryan, J. K., & Simchi-Levi, D. (2000). Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, *46*(3), 436–443. doi:10.1287/mnsc.46.3.436.12069

Cooper, E. L. M. G. J. T. (1997). Meshing multiple alliances. *Journal of Business Logistics*, *18*(1), 67.

Craig, C., Greene, W., & Douglas, S. (2005). Culture matters: Consumer acceptance of US films in foreign markets. *Journal of International Marketing*, *13*(4), 80–103. doi:10.1509/jimk.2005.13.4.80

Ding, C., Cheng, H. K., Duan, Y., & Jin, Y. (2017). The power of the "like" button: The impact of social media on box office. *Decision Support Systems*, *94*, 77–84. doi:10.1016/j.dss.2016.11.002

Elberse, A. (2007). The power of stars: Do star actors drive the success of movies? *Journal of Marketing*, *71*(4), 102–120. doi:10.1509/jmkg.71.4.102

Elberse, A., & Anand, B. (2007). The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy*, *19*(3-4), 319–343. doi:10.1016/j.infoecopol.2007.06.003

Escoffier, N., & McKelvey, B. (2015). The wisdom of crowds in the movie industry: Towards new so- lutions to reduce uncertainties. *International Journal of Arts Management*, *17*(2), 52.

Foutz, N. Z., & Jank, W. (2010). Research note: Rerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Science*, *29*(3), 568–579. doi:10.1287/mksc.1090.0542

Ghiassi, M., Lio, D., & Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. *Expert Systems with Applications*, *42*(6), 3176–3193. doi:10.1016/j.eswa.2014.11.022

Harrell, F. E. (2001). *Regression modeling strategies, with applications to linear models, survival analysis and logistic regression*. Springer. doi:10.1007/978-1-4757-3462-1

Helms, M. M., Ettkin, L. P., & Chapman, S. (2000). Supply chain forecasting: Collaborative forecasting supports supply chain management. *Business Process Management Journal*, *6*(5), 392–407. doi:10.1108/14637150010352408

Huang, D., Markovitch, A., & Strijnev, A. (2015). Exploring the small movie profitability puzzle. *Marketing Letters*, *26*(1), 43–55. doi:10.1007/s11002-013-9266-x

Hunter, Smith, & Singh. (2015). *Predicting Box Office from the Screenplay: An Empirical Model*. Academic Press.

Hur, M., Kang, P., & Cho, S. (2016). Box-office forecasting based on sentiments of movie reviews and independent subspace method. *Information Sciences*, *372*, 608–624. doi:10.1016/j.ins.2016.08.027

Karniouchina, E. V. (2011). Are virtual markets efficient predictors of new product success? the case of the Hollywood Stock Exchange. *Journal of Product Innovation Management*, *28*(4), 470–484. doi:10.1111/j.1540-5885.2011.00820.x

Latane & Nida. (1980). Social impact theory and group influence: A social engineering perspective. *Psychology of Group Influence*, 3-34.

Levmore, S. (2002). Simply efficient markets and the role of regulation: Lessons from the Iowa electronic markets and the Hollywood Stock Exchange. *J. Corp*, *28*, 589.

Marshall, P., Dockendorff, M., & Ibanez, S. (2013). A forecasting system for movie attendance. *Journal of Business Research*, *66*(10), 1800–1806. doi:10.1016/j.jbusres.2013.01.013

McCarthy, T. M., & Golicic, S. L. (2002). Implementing collaborative forecasting to improve supply chain performance. *International Journal of Physical Distribution & Logistics Management*, *32*(6), 431–454. doi:10.1108/09600030210437960

McKenzie, J. (2013). Predicting box office with and without markets: Do internet users know anything? *Information Economics and Policy*, *25*(2), 70–80. doi:10.1016/j.infoecopol.2013.05.001

Pennock, Lawrence, Giles, & Nielsen. (n.d.). *The power of play: Efficiency and forecast accuracy in web market games*. Academic Press.

Simonoff, J. S., & Sparrow, I. R. (2000). Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance*, *13*(3), 15–24. doi:10.1080/09332480.2000.10542216

Spann, M., & Skiera, B. (2003). Internet-based virtual stock markets for business forecasting. *Management Science*, *49*(10), 1310–1326. doi:10.1287/mnsc.49.10.1310.17314

U. S. C. Bureau. (n.d.). *Estimated domestic licensing revenue of the motion picture and video industry in the United States from 2010 to 2018 (in billion US dollars)*. Retrieved April 4, 2020. URL https://www.statista.com/statistics/185112/domestic-licensing-revenue-of-the-motion-picture-and-video-industry-since-2005

Kyongsei Sohn is associate professor in Department of Accounting, Economics and Finance at SUNY Brockport. His main areas of interest are applied microeconomics, applied econometrics and health economics. Before joining as a faculty member of SUNY Brockport, Dr. Sohn was a post-doctoral fellow at the University of Maryland, Baltimore.

John T. Gardner is a Professor Emeritus from SUNY Brockport. He has published in both Marketing and Logistics journals. He has been associated with SUNY Brockport since 1993.