Integration of AHP and Regression Analysis in Forecasting Attendance in a Movie Theater

Hossein Jamshidi

Professor of Operations Management, Management Marketing and Logistic Department

College of Business and Public Affairs, Alabama A&M University

E-mail: Hossein.jamshidi@aamu.edu

Larry McDaniel

Professor of Management, Management Marketing and Logistic Department

College of Business and Public Affairs, Alabama A&M University

E-mail: larry.mcdaniel@aamu.edu

Abstract

Forecasting the number of attendees in a motion picture has often been considered a wild guess since there are many factors or variables involved in forecasting the numbers. Among many possible variables the ones with the most significant impact considered in this study are; the day that movie is playing, the time of day, the ranking of the movie, the genre of movie, the length of time that the movie has been released and finally whether school is in or out. At first, the aim of this study is to compare all the variables by the decision maker based on the Analytical Hierarchy Process (AHP) and to rank them based on their importance according to the AHP. Once the variables are set, the regression is applied to forecast attendance in a movie theater. Multiple regression analysis is used based on a sample of 711 observations. Using SPSS statistical software, a model is developed to forecast the number of attendees and the model provides $R^2 = 0.760$, which is a strong predictor. Finally, the hypothesis test is conducted to verify the accuracy of the regression model with the actual data and even with $= 0.10$ the null hypothesis could not be rejected.

Keywords: Forecasting, AHP process, Regression, Decision Making,

1. Introduction

The movie industry is one of the fastest growing industries in the world. With a short life cycle, combined with the rapid and frequent introduction of a new movies, making a successful movie is very risky and managing and forecasting the success of a movie is extremely challenging. As a capital intensive industry, movie box office forecasting is very important for investment. The importance of movies and the need of society have made the motion picture industry one of the main sectors of today's global economy (Wang et al., 2010; Eliashberg, 2000). With many new movie releases around the world every year, global box office earnings increased to $34.7 billion in 2012, up to 6% over the previous year (Motion picture, 2012). Moreover, the film market is expected to experience constant growth up to 2017 (Price water house, 2013). Despite their inherent charms and promising prospects at the industry level, investing in motion pictures is still considered to be a high risk business for film makers, distributors, and even exhibitors, because of their alarming failure rates at the individual motion picture level (Shugan, 1995; Wang et al., 2010).

Forecasting the attendees in a movie theater is a crucial element for the industry's success and in particular to the local theater management. For planning and preparation purposes, the management would like to be able to foresee and guess the number of attendees in any particular showing within a multi-show theater. Many researchers have suggested numerous considerations or factors as important criteria for forecasting the attendees in a theater. (Litman, 1983) was one of the first to develop a multiple regression model in an attempt to predict the financial success of films. Litman's model provides evidence that the independent variables of production costs, critics' ratings, science fiction genre, major distributor, Christmas release, Academy Award nomination, and Academy Award win are all significant determinants for the success of a theatrical movie. (Litman & Kohl, 1989) and (Litman & Ahn, 1998) replicate and expand on the initial work by (Litman, 1983), while (Cameron, 1999) and (Dewenter & Westermann, 2005) investigate cinema demand in the UK and Germany. In a slightly different venue, (Terry, Butler, & De'Armond, 2004) examine the determinants of foreign box-office revenue for English language movies using U.S. domestic box-office revenue, action movies, children's movies, sequels, Academy Award wins, and the production budget. (Stimpert, Laux, Marino, & Gleason 2008) investigate the variables associated with the success of creative products and services by focusing specifically on movies. Their study offers a comprehensive examination of the factors that influence a movie's box-office success: quality, other product attributes, advertising expenditures, and the distribution pattern. Based on a review of the literature, the authors develop and test a model using a sample of 439 movies. All of these preceding models use cross-sectional data that fail to account for the dynamicity of demand and hence are unable to predict the life cycle of a particular movie. (Hand, 2002) proposes a time-series model for cinema admissions in the UK, and (Hand & Judge 2012) investigate whether Google Trends' search information can improve the forecasts of movie attendance. These models do not take into account either the information for particular movies or the dynamic of a movie's life cycle. These models are useful for the analysis of the industry as a whole but not for the managerial decisions at the theater level.

This study aims on forecasting attendees in a movie theater for any day of a week and at any

time of a day. Hence, we would like to have an answer to the question that how many audiences would attend any particular movie, shown on a particular day and at a particular time? For example, we found that Saturday evening is the most appealing time for movie goers, whereas week days' matinee is the least desired time to go to a theater. The expected attendees for a movie will influence the size of screening room where it will be shown. Moreover, the demand for some movies may be so large that it could be double or triple booked. Also, there are different genres of movies, and this has implications for their scheduling. Children's movies should, preferably, be shown at times when children are free from school (weekends and Wednesday afternoons), and will not be shown during evenings. Accurate forecasts benefit the entire value chain in many ways, from enabling more informed investment decisions to planning better marketing and entry time strategies. However, having accurate forecasts in the early stages of a movie is very difficult owing to a distinctive characteristic of motion pictures. Because a movie is an experiential product, a consumer has limited access to information about it before he or she actually watches it.

At first, a regression model is developed with inclusion of all variables under consideration in this research. The result of the regression model is analyzed and then all the variables are ranked by the decision maker based on AHP process to see if any variable could be omitted from the analysis. If any variable is omitted, then another regression model will be developed based on the reduced model and the results will be compared with the original model. According to Decision Support System Resources, the definition of the Analytical Hierarchy Process (AHP) is an approach to decision making that involves structuring multiple choice criteria into a hierarchy, assessing the relative importance of these criteria, comparing alternatives for each criterion, and determining an overall ranking of the alternatives. (Saaty, 1980), the founder of the AHP, claims that the AHP is "natural to our intuition and general thinking," which combines logic and intuition and takes advantage of our ability to rank choices. The AHP asks respondents to rank them by using pair-wise comparisons and assign a number, representing weights, to each pair comparison. The numbers will generate a matrix giving the decision maker an organizational tool to attack the larger problem. Additionally, AHP helps capture both subjective and objective evaluation measures, providing a useful mechanism for checking the consistency of the evaluation measures (Parcom, 2007). When the decision cycle involves taking into account a variety of multiple criteria, AHP splits the overall problem to solve into many evaluations of lesser importance, while keeping at the same time their part in the global decision.

2. Methodology

This study contributes to a better understanding and prediction of the attendees for a movie in multi showing theater. The study offers two methodologies for forecasting attendance in a movie theater. The AHP is first applied to indicate what variables are most important to the decision maker and then the multiple regression analysis is developed and applied. A sample of 711 movie goers is obtained over a period of almost six months for audiences attending a movie theater. Many variables could be included in forecasting attendees in any given time and day of a week. Among many possible variables this study aims on the variables of: DAY of the week, SHOW time of a day, RANK of the movie, WEEKs a movie is on screen,

SCHOOL in/out and GENRE. The Show time is set at 5 levels of 10-12, 12-3, 3-6, 6-9 and 9-11. The Day of week is having 4 levels of Friday, Saturday, Sunday and Monday through Thursday. Monday through Thursday is considered as week days and lumped in to one category. Rank of a movie is considered at 3 levels of very good (hit) film, good film and the average (OK) film. School is set at two levels of in and out. Finally, Genre is set at 6 levels of Family, Comedy, Action, Suspense/horror, 3-D, Kids and Else.

3. Results

Initially total of 24 variables are considered in this study. For the initial model Show time of a day is appeared at 5 levels, Day of a week at 4 levels, Rank of a movie at 3 levels, Week or release time is appeared at 4 levels, School in/out at 2 levels and finally Genre at 6 levels. The variables are summarized below.

Show =	X_{1S}	*where*	*S = 1 to 5*
Day =	X_{2D}	*where*	*D = 1 to 4*
Rank =	X_{3R}	*where*	*R = 1 to 3*
Week =	X_{4W}	*where*	*W = 1 to 4*
School-in/out =	X_{5I}	*where*	*I = 1 to 2*
Genre =	X_{6G}	*where*	*G = 1 to 6*
Attendance =	*Y*		

A random sample of 711 observations over a period of almost 6 months were collected and using SPSS and a regression model is developed for forecasting the number of attendees in a theater. Table 1 represent the Model Summary and ANOVA table is given below. The $R^2 = 0.76$ with standard error of 7.732.

Table1. ANOVA Table with all the Variables

Model	SS	df	MS Error	F	significant
1					
	129918.714	22	5905.396	98.785	0.00000
Residual	41128.805	688	59.78		
Total	171047.519		710		

The developed regression model is as follow;

Y = 34.385-17.13X₁₁-13.37X₁₂-7.23X₁₃+4.28X₁₄+8.8X₁₅+7.67X₂₁+8.57X₂₂+9.54X₂₃-11.72X₃₂-25.00X₃₃+7.58X₄₁-6.68X₄₃-12.77X₄₄-.99X₅₁-3.89X₅₂+0.24X₆₁+1.53X₆₂+1.20X₆₃+0.36X₆₄+2.78X₆₅+1.53X₆₆+0.75X₆₇

$$Y = 34.385 - 17.13X_{11} - 13.37X_{12} - 7.23X_{13} + 4.28X_{14} + 8.8X_{15} + 7.67X_{21} + 8.57X_{22} + 9.54X_{23} - 11.72X_{32} - 25.00X_{33} + 7.58X_{41} - 6.68X_{43} - 12.77X_{44} - .99X_{51} - 3.89X_{52} + 0.24X_{61} + 1.53X_{62} + 1.20X_{63} + 0.36X_{64} + 2.78X_{65} + 1.53X_{66} + 0.75X_{67}$$

The result of Table 2 reveals that at α = .05, the variables of Friday, Saturday, Sunday, good movie, OK movie, 1st, 3rd and 4th weeks are highly significant and are a very good predictor of the variable of attendance. When looking at the model more closely, it is observed that the

variables of time, (11-1) and (1-3), are at marginal level and an interesting observation is that neither the variables of School in/out nor the Genre had any significant influence in performance of the model. After this finding, the AHP process is applied to prioritize the goals. AHP helps capture both subjective and objective evaluation measures, providing a useful mechanism for checking the consistency of the evaluation measures and alternatives suggested by the decision maker.

Table 2. Coefficients of Model 1 for All the Variables

Model-1	B	Std. Error	t	Significant
Constant	34.385	8.902	3.863	0.00
Time (11-1)	-17.130	7.984	-0.215	0.032
Time (1-3)	-13.369	7.908	-0.351	0.091
Time (3-6)	-7.269	7.896	-0.208	0.358
Time (6-9)	4.276	7.901	0.541	0.589
Time (9-11)	8.801	7.901	1.114	0.266
Friday	7.666	1.110	6.907	0.000
Saturday	8.567	1.045	8.199	0.000
Sunday	9.542	0.958	9.963	0.000
Good	-11.717	0.922	-12.711	0.000
OK	-25.008	0.978	-25.570	0.000
1st week	7.577	0.722	10.493	0.000
3rd week	-6.684	0.917	-7.290	0.000
4th week	-12.771	1.074	-11.891	0.000
School out	-0.990	2.890	-0.343	0.732
School in	-3.892	2.869	-1.357	0.175
Family	0.243	2.890	0.084	0.933
Comedy	1.527	2.865	0.533	0.594
Action	1.204	2.862	0.420	0.674
Suspense	0.364	2.938	0.124	0.902
3-D	2.781	2.937	0.947	0.344
Kids	1.502	3.035	0.495	0.621
Else	0.752	3.095	0.243	0.808

Table 3 is designed to elicit and assess information on preferences for the variables under consideration. The variables of DAY of week, SHOW time, RANK of movie, WEEKS on screen, SCHOOL in/out and GENRE are considered in this study. The variables are compared by decision maker based on paired wise comparisons. The entries of this table are obtained as a result of pair wise comparison. For example, the decision maker has assigned a weight of 5 to 1, comparing Day of the week to School in/out, meaning the day of the week is 5 times as important as school is in or out.   Likewise, school is 1/5 as important as day of the week. Also, for example, rank of a movie is 7 times as important as genre.

Table 3.   Priority Assignment of the Variables

	Day	Show	Rank	Week	School	Genre
Day	1	2	0.25	0.5	5	7
Show	0.5	1	0.5	1	4	5
Rank	4	2	1	2	5	7
Week	2	1	0.5	1	3	5
School	1/5	¼	1/5	1/3	1	1
Genre	1/7	1/5	1/7	1/5	1	1
Total	7.84	6.45	2.59	5.03	19	26

Table 4 is developed based on percentages of preferences of the decision maker and this table shows the average percentages for the priorities.   It shows that the RANK of a movie is the most important factor to forecast attendance.   The DAY of a week followed number of WEEK movie is on screen are also important variables.   The least important variable is the GENRE followed by the SCHOOL is in or out. The result of AHP process is used as an input to develop a reduced regression model. This result reinforces the finding of the regression model in which the variables of Scholl in/out and Genre were the lowest ranked among all the variables.

Table 4. Priority Assignment of the Variables

	Day	Show	Rank	Week	School	Genre	Average %
Day	1/7.84=0.127	2/6.45= 0.310	0.25/2.59=0.097	0.5/5.03=0.099	5/19=0.263	7/26=0.269	0.194
Show	0.5/7.84=0.064	1/6.45= 0.155	0.5/2.59=0.193	1/5.03=0.199	4/19=0.210	5/26=0.192	0.168
Rank	4/7.84=0.510	2/6.45=0.310	1/2.59=0.386	2/5.03=0.398	5/19=0.263	7/26=0.269	0.356
Week	2/7.84=0.255	1/6.45=0.155	0.5/2.59=0.193	1/5.03=0.199	3/19=0.158	5/26=0.192	0.192
School	1/5/7.84=0.026	¼/6.45=0.039	1/5/2.59=0.077	1/3/5.03=0.066	1/19=0.053	1/26=0.038	0.049
Genre	1/7/7.84=0.018	1/5/6.45=0.031	1/7/2.59=0.055	1/5/5.03=0.039	1/19=0.053	1/26=0.038	0.032
Total							0.992

Hence, model 2 is developed based on the removal of the variables of GENRE and School in and out. The variable of GENRE includes family, comedy, action, suspense, 3-D, kids and else.   The idea is that if a movie is good or bad it does not matter what kind of movie it is; the attendees will not be interested in that movie.   Also, whether the school is in or the school is out, having no effect on movie goers during the weekend and if they like a movie they will watch it anyway. After exclusion of the nine variables which were not significant in forecasting the attendees in a theater the following tables were generated. Model 2 is Based on the result of the reduced regression model and the following regression model is developed.

Y =

30.373-14.820X_{11}-10.820X_{12}-4.724X_{13}+6.825X_{14}+11.352X_{15}+7.573X_{21}+8.830X_{22}+9.464X_{23}-11.532X_{32}-

      25.229X_{33}+7.270X_{41}-6.413X_{43}-12.866.

The R^2 = 0.75 for model 2, which is a good indication of performance of model 2. Moreover, comparing the R^2 of model 1 and 2 indicates and reinforces that the variable of genre was not effective in predicting the attendees in a theater and the model performance was not changed significantly. Table 5 represents the ANOVA table for the reduced model.

Table 5. ANOVA Table with the Reduced Variables

Model 2	SS	df	MS Error	F	significant
	128220.663	13	9863.128	160.521	0.00000
Residual	42826.856	697	61.445		
Total	171047.519	710			

The result of Table 6 reveals that at   = .05, almost all the variables are significant in forecasting the attendees. Surprisingly the variables which were not significant in the new reduced model are time 11-1, time 1-3, time 3-6, time 6-9 and time 9-11. The variables which were highly significant are the days Friday, Saturday, Sunday, good, OK movies, and 1st, 3rd and 4th week on the screen showing.

Table 6. Coefficients of Model 2 with the Reduced Variables

Model-2	B	Std. Error	t	Significant
Constant	30.373	8.012	3.791	0.000
Time (11-1)	-14.820	8.062	-1.838	0.066
Time (1-3)	-10.820	7.977	-1.356	0.175
Time (3-6)	-4.724	7.964	-0.593	0.553
Time (6-9)	6.825	7.970	0.856	0.392
Time (9-11)	11.352	7.969	1.424	0.155
Friday	7.573	1.098	6.898	0.000
Saturday	8.830	1.048	8.424	0.000
Sunday	9.464	0.942	10.044	0.000
Good	-11.532	0.908	-12.706	0.000
OK	-25.229	0.965	-26.147	0.000
1st week	7.270	0.719	10.111	0.000
3rd week	-6.413	0.902	-7.107	0.000
4th week	-12.866	1.072	-12.000	0.000

Finally, the performance of model 2 is evaluated by comparing the result of the regression model with the actual number of attendees in the theater. The hypothesis test was conducted to measure if there is any difference in the Actual number of attendees versus the number of attendees obtained by the regression model as follow:

Ho: *there is no difference between the result of the regression model and the actual number*

H1: *There is*

A sample of n=20 observation is generated randomly, based on different situations, by regression model 2 and the results are compared with the actual attendees under the same condition or situations as to the regression model. For example; the number of attendees for 6-9 PM on Friday where the movie is rated Good and is on first week of showing is 40.51 however the actual number is 37.   Table 7 compares the results for the Actual numbers and the forecasted one by the regression model.

A t-test for the matched sample is conducted as follow;

$$t = \frac{\overline{d} - \mu d}{\frac{Sd}{\sqrt{n}}}$$

 Therefore;

$$t = \frac{1.2935 - 0}{\frac{5.9202}{\sqrt{20}}} = 0.9771$$

For n=20 with 19 degrees of freedom, the p-value reveals that the null hypothesis cannot be rejected even with   = 0.10. Therefore, we conclude there is no difference in responses of the Actual data versus the Forecasted data obtained from the regression model.

Table 7. Comparing the Actual result with Forecasted result

Observation	Actual	Forecast	Difference
1	25.80	20.12	5.68
2	28.01	21.25	6.76
3	50.50	42.26	8.24
4	9.50	13.25	-5.61
5	22.25	33.86	-11.61
6	37.00	40.51	-3.51
7	48.00	45.04	2.96
8	30.80	24.13	6.67
9	34.00	30.22	3.78
10	14.4	6.84	7.56
11	44.80	46.92	-2.12
12	32.67	37.23	-4.56
13	15.60	19.5	-3.9
14	53.80	45.25	8.55
15	11.00	8.85	2.15
16	26.50	34.5	-8.00
17	25.45	18.75	6.70
18	27.75	28.3	-0.55
19	30.67	33.5	-2.83
20	20.45	16.6	3.85
Mean			1.2935
Standard dev.			5.9202

4. Conclusion

In this paper, we attempted to develop a model to forecast the number of attendees in a theater by considering a wide range of variables. The main objective of this paper is to develop a model that uses historical data to forecast movie attendance. Among many possible variables this study aims on the variables of: DAY of the week at 4 levels, SHOW time of a day at 5 levels, RANK of the movie at 3 levels, WEEKs a movie has been on screen at 4 levels, SCHOOL in/out at 2 levels and finally and GENRE at 6 levels. Total of 711 data were captured over a period of almost 6 months. A regression model was developed based on SPSS statistical package. The result of model indicated an $R^2 = 0.76$ which is a good indicator of the model. Among all 24 variables the variables of GENRE at 6 levels and SCHOOL in/out at 2 levels were not significant in forecasting attendees in a movie theater. This finding was reinforced by the finding of AHP process. After removing variables from the model and running the regression model again the new reduced regression model generated $R^2 = 0.75$ which is almost the same as the $R^2 = 0.76$ of the original model. This will justify the exclusion of the insignificant variables from the model. In addition, a hypothesis test is conducted to validate the result of the regression model with the actual data, using the t-test for matched sample, the null hypothesis could not be rejected even with = 0.10.

Despite the interesting experimental results, there are some limitations to the current work which provide future research directions. First the data collected was not as large as we wished to collect because the time was an issue and we had limited time. Also the period of six months was not long enough to capture the effect of seasonal, holidays and monthly behavior. For more accurate market behavior, the length of at least a year should be consider to capture all year seasonal behavior.

References

Cameron, S. (1999). Rational addiction and the demand for cinema. *Applied Economic Letters, 6,* pp. 617-620. https://doi.org/10.1080/135048599352736

Decenter, R., & Wasserman, M. (2005). Cinema demand in Germany. *Journal of Cultural Economics*, *29*, 213-231. https://doi.org/10.1007/s10824-005-6421-0

Eliashberg, J., Jonker, JJ., Sawhney, M.S. & Wierenga, B. (2000). Motivated: An implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Science*, 19(3), 226-243. https://doi.org/10.1287/mksc.19.3.226.11796

Hand, C. (2002), The distribution and predictability of cinema admissions. *Journal of Cultural Economics*, *26*, 53-64. https://doi.org/10.1023/A:1013389211323

Hand, C. & Judge, G. (2012), Searching for the picture: Forecasting UK cinema admissions using Google Trends data. *Applied Economics Letters*, *19*, 1051-1055. https://doi.org/10.1080/13504851.2011.613744

Litman, B.R. (1983, spring), Predicting success of theatrical movies: An empirical study; *The Journal of Popular Culture*, *16,* 159-175. https://doi.org/10.1111/j.0022-3840.1983.1604_159.x

Litman, B.R. & Kohl, A. (1989). Predicting financial success of motion pictures: The '80s experience. *The Journal of Media Economics*, *2*(1), 35-50. https://doi.org/10.1080/08997768909358184

Litman, B.R., & Ahn, H. (1998). *The Motion Picture Mega-industry,* Allyn & Bacon Publishing, Inc., Boston, MA.

Parcom, P. (2007). *Developing your decision making skills.* New Delhi, India: Lotus Press.

Price water house Coopers, (2013). "PwC global media outlook: 2013–2017," New York.

Saaty, L. T. (1980). *The analytical hierarchy process.* New York: McGraw Hill. https://doi.org/10.21236/ADA214804

Shugan, S. (1995). *Service Strategy,* Richard D. Irwn, Homewood, Il, USA.

Stimpert, J.L., Laux, J.A., Marino, C., & Gleason, G. (2008). Factors influencing motion picture success: mpirical review and update, Journal *of Business & Economics Research*, *6*, 39-52. https://doi.org/10.19030/jber.v6i11.2488

Terry N., Butler, M., & De'Armond, D.A. (2004). The economic impact of movie critics on box office performance. *Academy of Marketing Studies Journal*, *8*(1), 61-73.

Wang, F., Cai, R., & Huang, M. (2010). Forecasting movie-going behavior based on online pre-release WOM and opening strength. Proceedings of the 2nd International *Workshop on Intelligent Systems and Applications*, (ISA'10), Wuhan, China. https://doi.org/10.1109/IWISA.2010.5473531

Wang, F., Zhang, Y., Li, X., & Zhu, H. (2010) Why do moviegoers go to the theater? the role of prerelease media publicity and online word of mouth in driving movie going behavior. *Journal of Interactive Advertising*, *11*(1), 50-62. https://doi.org/10.1080/15252019.2010.10722177