Prediction Model of Box Office Based on Arbitrage Pricing Theory: An Empirical Analysis from China

Wang Qingshi, Li Naiqian, Hashmat Ali*

Dongbei University of Finance and Economics, Dalian, Liaoning, China. *Email: hashmatali11@qq.com

ABSTRACT

According to characteristics of revenue and risks involved in film investments, this paper expounds that movies can also be treated as a common tradable asset in capital markets. Consequently, this study employs the famous Arbitrage Pricing Theory (APT) from Capital Asset Pricing Model as a necessary and efficient theoretical explanation, in order to describe the application of multi-factor linear regression model used in assessing movie investment returns in real life. Above that, the paper not only uses Linear Regression with Multiple Variables and its theory by Litman and Kohl (1989) for reference, but also build a model for explanation and prediction about Chinese movie box office combined with the research experience in predicting movie box office at home and abroad. 219 movies released in Mainland China from 2008 to 2017 were selected as samples for empirical testing and analysis. According to the empirical model, this study predicted the 4 movies released in 2018, satisfactory outcomes were obtained.

Keywords: Movie Box Office, Prediction Model, Arbitrage Pricing Theory, Empirical Analysis
JEL Classifications: C1

1. INTRODUCTION

Film has always been a high-risk and high-return investment project. The main revenue of the film projects comes from the revenue generated by the right of broadcast. Especially in China, 90% of film project income is box office income (Han, 2010). Consequently, the prediction of film project income in China is mainly the prediction of box office income. If the box office income can be predicted scientifically and effectively in advance according to the theoretical model and empirical information, it is of great practical significance to guide investment more rationally and reduce the uncertainty risk of investors' investment returns.

For a long time, box office prediction models, even the classic Litman multiple linear regression model, have been highly dependent on practical experience and lack of theoretical support. The research on film box office can be traced back to the 1940s, and the research on film box office through modeling has a nearly 30-years history. At present, the most classic and mature box office

prediction model is a series of empirical studies on box office prediction based on the multi-factor linear regression model set up by Litman in 1989. However, neither the classic Litman movie box office prediction model nor a series of subsequent empirical studies based on it have elaborated the theoretical inevitability of the model itself. This study will treat film products as common capital assets, because of the film proceeds is a variety of risk factors impact the risk premium. Thus, the multi-factor pricing model system of the capital asset pricing theory could be used. So as to consolidate the movie box office revenue forecasts and theoretical basis of pricing model, it will further promote the film asset pricing analysis and study of the breadth and depth.

2. LITERATURE REVIEW

Litman's and Kohl (1989) paper, "Predicting Financial Success of Motion Pictures: The 80's Experience," established the research method of predicting box office by using a multivariate linear

regression model. The regression analysis method is still the mainstream research method in the academic world. Litman and Kohl (1989) took the rental income of films as the dependent variable because it was hard to accurately count the box office of films at that time. He divided the factors that affected box office into three categories: creative sphere, release pattern and marketing effort. The creative sphere category has seven sub-categories: type of story (Genre), the MPAA ratings, star actors and top directors, production budget and critics' rating; There are four sub-categories in the release pattern category: distributor, release mode, release schedule and market participation; There are two sub-categories of marketing effort: the marketing ability of the distributor and whether the film had won any awards. There were five genres of films in his study: comedy, drama, science fiction, action and music. The MPAA ratings were classified into PG, G and R. The release dates were divided into Christmas, Summer and Easter. Litman used 14 influencing factors as independent variables in the prediction model.

Based on Litman model, Scott (1994) selected films released in North America from 1987 to 1989 as samples and conducted an empirical study on box office prediction. The dependent variables in the study were domestic rentals and the length of run of each film. Based on Litman's study, Scott further refined the independent variables, and the independent variable factors increased from 14 to 22. The biggest innovation of Scott's model was to introduce the concept of market concentration into the model.

Chang and Ki (2005) criticized the logic flaw of introducing some factors into the previous box office prediction model. Because it is generally impossible to acknowledge whether a film would win any award before its release, the variable "Award" was adopted in their study. In addition, they referred to the research experience of Broadway stage play box office (Reddy et al., 1998) and introduced the concept of brand into film research: as an extension of the concept of brand, the dummy variable "Sequel" was using to improve the prediction model of film box office. The most important innovation in their research was measuring stars or directors quantitatively from horizontal aspect and vertical aspect instead of a qualitative analysis using dummy variables. The horizontal aspect is the box office performance of the most recent movie an actor appeared in. The horizontal aspect is the total number of movies the actor has appeared in during his or her career. (Jianling, 2009).

The research on film box office in China started relatively late, and there are few Chinese literatures in relevant fields. Most of the research in China are empirical researches based on existing models. Yiyi (2013) from Tsinghua University conducted an empirical analysis based on Litman model with 74 domestic films which box office exceeded 100 million Yuan from 2002 to 2012 as samples. In the same year, Dan (2013) from Communication University of China took 3D films released in China as the research object for prediction and analysis. Jinhui and Siyu (2015) from Communication University of China selected 196 films released in China from 2010 to 2014 as samples for empirical research, and they introduced 3D picture quality as a variable into theirs model. However, the empirical results showed that the variable was not significance. Jinhui et al. (2016) selected 50 typical domestic movie to test the various factors influencing the film at the box office model, the China's domestic

movie box office prediction model had been concluded. In the analysis, the director and actors, movie's special effects, the 1st week of box office is significantly correlated with total box office. Critic's ratings, distribution companies, releasing schedule are no significant in correlation. Therefore, they came to the conclusion that it was the film quality determine the box office. Consequently, the film production should focus on the quality of the film.

3. ARBITRAGE PRICING THEORY (APT) AND THE PREDICTION MODEL OF CHINESE FILM BOX OFFICE

Throughout the research history of film box office prediction, scholars unconsciously regarded film as a special commodity for research (Chang and Ki, 2005). This study believes that film products also have many characteristics consistent with the characteristics of capital assets: First, Profitability. The copyright transaction and box office earnings of film works can generate profits for the investors of film projects. Second, Time Horizon. Revenue from films is not instantaneous, and most film projects take two to three years to generate revenue from the first investment. Third, Tradability. As intangible assets, film and television works can be divided and traded in the form of shares at any time. Fourth, the characteristics of high return and high risk. The film return is affected by a series of risk factors such as actor, director, genre and release schedule, and the film return is the risk premium of these various risk factors. To sum up, film products can be regarded as an asset that can be traded in the capital market. Therefore, factors influencing the earnings of film products can be analyzed according to APT, so as to qualitatively and quantitatively identify the relationship between film earnings and influencing factors. In China, 90% of the film revenue comes from the box office, so the prediction of the investment income of Chinese films is mainly the prediction of the box office income of Chinese films.

3.1. APT and Multiple Linear-Factor Model

Ross (1976) brought an asset pricing theory: APT. APT is based on three assumptions:(1) investors have the same investment philosophy; (2) investors pursue the maximization of returns at a certain level of risk, or the minimization of risks at a certain level of return; (3) investors are in the complete market. According to the hypothesis, this theory believes that in an effective capital market, if the market does not reach an equilibrium state, then the opportunity of risk-free arbitrage will appear. Because of the existence of arbitrage opportunities, it is profitable for investors who avoid risks. Therefore, once such opportunities are found, rational investors will make use of them until the arbitrage opportunities disappear, while the market reaches equilibrium. APT is actually derived by mathematical method under the assumption of market equilibrium without arbitrage. Hence, the theory is expressed by a rigorous multivariate linear regression model, which indicates that the change of return on any asset in the capital market is influenced and determined in a linearization by multiple factors in the market. The general model of APT is expressed as follows:

$$r_i = a_i + \beta_{i1}F_1 + \beta_{i2}F_2 + \beta_{i3}F_3 + \cdots + + \beta_{ik}F_k + \delta_i$$

Where r_i is the return of the i-asset, β_{ik} is the coefficient of influence of the risk factor F_k on the return on the asset, F_k is the kth factor of the return on the asset, and δ_i is the random disturbance term. The model is actually the theoretical root cause of the multiple linear regression models used in academic box office revenue or earnings forecasts.

3.2. Prediction Model of Chinese Film Box Office

In this study, the film is treated as a general trading capital assets, because the box office is the actual form of movies earnings, so according to the APT of multivariate linear regression modeling technology and how to construct the variable to analyze the influence factors of Chinese movie earnings, combining with the experience of Chinese and foreign movie box office forecasting, Chinese movie box office forecasting model is constructed.

The number and type of risk factors which influencing asset income did not be given by APT, so in the actual modeling, the factors must be determined for the basic structure of multi-factor model. Commonly, exploratory factor analysis and the designated factor analysis are typically methodologies. It is to explore the basic structure of the observed data under the assumption that the public risk factors affecting asset returns are not observable. Exploratory factor analysis studies the internal dependence among many observational variables. A few hypothetical factors are used to represent the basic data structure of arbitrage pricing model, so as to achieve the purpose of concentrating observation data with minimum information loss. Designated factor analysis assumes that public risk factors are observable. The number and type of public risk factors are determined according to economic theory and subjective judgment of investment experts. Multivariate regression analysis in statistics is used to estimate the multi-factor model of arbitrage pricing. The estimated regression coefficient is used as the sensitivity of asset returns to public risk factors (Su Ping, 2005).

In this study, according to the characteristics of the film itself, the method of designated factor analysis is used to determine the factors in the arbitrage pricing model and adds the factors as variables into the model which shown in Table 1.

3.2.1. Dependent variable
In this paper, the dependent variables are of the total box office and the 1st week box office which released in China mainland. Since China's film cost data cannot be accurately obtained, it is impossible to calculate the film return rate. However, under the premise of fixed cost, returns and return rate are linearly related. Therefore, in the empirical study of this paper, we will use box office income to replace return rate for regression analysis.

3.2.2. Independent variable
For the prediction of Chinese box office, the author combined the research results of previous scholars and the characteristics of Chinese film market, and selected the director, actor, actress, release schedule, film type and critics rating as the influencing factors for the study.

3.2.3. Actor, actress and director
This paper adopts the ideas of Chang and Ki which any actor, actress and director can be quantified with general data in both horizontal and vertical separation. Horizontal data of actors, actresses and directors is the box office income of whom participated in a film recently. Vertical indicators are the number of films whom had participated in before.

3.2.4. Type (Genre) of story
This variable examines whether the type of film influences the box office revenue of the film. In this study, film types were divided into action, romance, war, comedy, drama, science fiction and horror. They are all dummy variables when they participated in the model. Multiple-element type film is a popular form in recent years. In order to explore whether multiple-element presentation at

Table 1: Variables and definitions

Variables	Abbreviations	Definitions
Total Box Office	BO	Total box office of a film released in Mainland China (10 million)
First Week Box Office	FBO	First week box office of a film released in Mainland China (10 million)
Actors	M1	Box office of the most recent movie the actor appeared in (10 million)
	M2	The total number of movies the actor has appeared in during his career
Actress	F1	Box office of the most recent movie the actress appeared in (10 million)
	F2	The total number of movies the actor has appeared in during her career
Directors	DIR1	The box office of the directors most recent movie (10 million)
	DIR2	The total number of movies the director has directed during his/her career
Genre	ACT	Action, dummy variable
	ROMANCE	Romance, dummy variable
	WAR	war, dummy variable
	COMEDY	comedy, dummy variable
	DRAMA	Drama, dummy variable
	SCI	Science fiction. Dummy variable
	HOR	Horror, dummy variable
	MULIT	A Movie contain the number of the above types, ordinal level variable, from 1 to 7
Critics rating	RATE	Critics reviews rating. From 1 to 10
Release date	SUMMER	Films first released date during 1st June to 31st July
	NATIONAL	Films first released date during 20th September to 30th October
	NEWYEAR	Films first released date during 15th December to 20th February

the same time will have a positive impact on box office, the author designed an ordinal level variable in the model, which indicates the number of containing types in a film. The value of the ordinal level variable is 1-7.

3.2.5. Critics ratings

Critics ratings is an intuitive manifestation of moviegoers' recognition of a movie. Generally, the higher the rating, the more people will be attracted to watch the movie thus raising the box office revenue of the movie.

3.2.6. The release schedule

The release schedule is usually accompanied by holidays. Audiences will have more leisure time to go to the cinema, which will affect the box office revenue positively. Due to the different conditions between China and other countries, the schedules cannot be directly adopted in this paper. For example, the Easter and Christmas is western countries festival, and do not conform to China's national conditions. So this study divides the schedules in China into Summer period, National Day Festival and New Year Vacation according to the actual situation. This paper defines the film released on June 1 to July 31 as the summer films. Films released from September 20 to October 30 are defined as National Day films. Films released on December 15 solstice and February 20 are defined as New Year films. All scheduling factors are studied in the form of dummy variables in the model.

3.3. Prediction Model of Chinese Film Box Office

In summary, according to the APT model and Litman's and Kohl (1989) multivariate movie box office income linear regression model, combined with the above-mentioned influence variables in the Chinese film market, the specific form of the Chinese box office predation model is:

$$
\begin{aligned}
BO_i = \beta_0 &+ \beta_1 M1_i + \beta_2 M2_i + \beta_3 F1_i + \beta_4 F2_i + \beta_5 DIR1_i + \\
&\beta_6 DIR2_i + \beta_7 ACT_i + \beta_8 WAR_i + \beta_9 LOVE_i + \\
&\beta_{10} COMEDY_i + \beta_{11} DRAMA_i + \beta_{12} SCI_i + \beta_{13} HOR_i + \\
&\beta_{14} MULTI_i + \beta_{15} RATE_i + \beta_{16} SUMMER_i + \\
&\beta_{17} NATIONAL_i + \beta_{18} NEWYEAR_i + \varepsilon_i
\end{aligned}
$$

The 1st week box office predation model is:

$$
\begin{aligned}
FBO_i = \beta_0 &+ \beta_1 M1_i + \beta_2 M2_i + \beta_3 F1_i + \beta_4 F2_i + \beta_5 DIR1_i + \\
&\beta_6 DIR2_i + \beta_7 ACT_i + \beta_8 WAR_i + \beta_9 LOVE_i + \\
&\beta_{10} COMEDY_i + \beta_{11} DRAMA_i + \beta_{12} SCI_i + \beta_{13} HOR_i + \\
&\beta_{14} MULTI_i + \beta_{15} RATE_i + \beta_{16} SUMMER_i + \\
&\beta_{17} NATIONAL_i + \beta_{18} NEWYEAR_i + \varepsilon_i
\end{aligned}
$$

4. EMPIRICAL ANALYSIS OF BOX OFFICE PREDICTION MODEL IN CHINA

4.1. Data Sources

This paper studies the factors that influence the box office of the Chinese mainland regional film market. Therefore, we selected films released in mainland of China as samples. However, due to limited data sources, we could only select the top 25 films released in mainland of China from 2008 to 2017 as samples from Entgroup database (http://www.cbooo.cn/). Since the production mode of animated films is different from general films', animated films are excluded from the samples. After processing, the author will take 219 films' box office and relevant data as samples for empirical study. Due to the time span, the author regards 2008 as the basic year (CPI=100), and adjusts the box office of different years according to the consumption index of each year. Critics' ratings come from the data sources (https://movie.douban.com/) on October 1, 2018 douban score of each film. The samples are cross-section data. The processed cross-section data are input into STATA for regression analysis, and the correlation between independent variables and dependent variables is studied.

4.2. Empirical Results

Correlation detection was performed on all explanatory variable data in the model. The Pearson correlation coefficient is shown in Table 2:

It can be seen from this table that all the independent variable correlation coefficients are below 0.5, and most of them are below 0.2. Visually observation of the two models would not lead to the possibility of collinearity due to independent variables. However, since the independent variables in the model have a branch of multiple dummy variables, the simple Pearson correlation statistic detection may be invalid. Therefore, we introduce the Variance Inflation Factor (VIF) to detect each independent variable. The average VIF of the 18 independent variables is 1.47, and the largest VIF of the variable ACT is 2.65, which is much <10. Thus, the collinearity problem of the models is under a high degree of tolerance.

The author performs a white test on the total box office model and the 1st week box office model. The white test statistical values are 127.39 and 146.26, and the partial probability P values are 0.7715 and 0.3414. The statistical results show that the null hypothesis "the model does not have heteroscedasticity" cannot be rejected at the 10% significance level. However, in order to avoid the possibility of heteroscedasticity, we still use the robust estimation method for the regression model. The regression results are as follows:

The total box office predication estimated model is:

$$
\begin{aligned}
\hat{B}O_i = &\underset{(0.75)}{15.6098} + \underset{(2.95)^{***}}{0.5838\,M1_i} + \underset{(2.16)^{**}}{0.0146\,M2_i} + \underset{(1.27)}{0.0593\,F1_i} - \\
&\underset{(-2.03)^{**}}{0.4891\,F2_i} - \underset{(-0.07)}{0.0074\,DIR1_i} + \underset{(0.02)}{0.0073\,DIR2_i} - \\
&\underset{(-1.11)}{11.1399\,ACT_i} - \underset{(-2.79)^{***}}{34.3842\,WAR_i} - \underset{(-2.40)^{**}}{22.5884\,LOVE_i} + \\
&\underset{(0.53)}{7.0469\,COMEDY_i} - \underset{(-0.84)}{11.2524\,DRAMA_i} - \underset{(-0.28)}{3.0240\,SCI_i} - \\
&\underset{(-4.86)^{***}}{45.9417\,HOR_i} - \underset{(-1.42)}{6.9229\,MULTI_i} + \underset{(2.18)^{**}}{5.4899\,RATE_i} + \\
&\underset{(1.99)^{**}}{26.9775\,SUMMER_i} + \underset{(0.85)}{6.7490\,NATIONAL_i} + \\
&\underset{(1.71)^{*}}{13.0083\,NEWYEAR_i} \qquad\qquad R^2 = 0.1795
\end{aligned}
$$

Table 2: Pearson correlation coefficient

	M1	M2	F1	F2	DIR1	DIR2	ACT	romance	WAR	COMEDY	DRAMA	SCI	HOR	MULIT	RATE	SUMMER	NATIONAL
M2	0.0044																
F1	-0.0251	-0.0478															
F2	0.0466	-0.0597	0.2973														
DIR1	0.2161	-0.0181	-0.0009	0.1063													
DIR2	-0.0822	0.3666	-0.0351	-0.1345	-0.0539												
ACT	0.0998	-0.0297	-0.0221	-0.0386	0.0498	-0.0951											
romance	-0.0494	-0.0047	-0.0502	-0.0889	0.0769	0.1517	-0.1509										
WAR	-0.0100	0.1377	0.1623	0.0895	-0.0285	0.1327	-0.2791	-0.0793									
COMEDY	0.1563	0.0172	0.0564	0.0426	0.0641	-0.0543	-0.2791	-0.0793	-0.1466								
DRAMA	-0.1029	-0.0305	-0.0120	0.1012	-0.0228	-0.0022	-0.1420	-0.0403	-0.0746	-0.0746							
SCI	-0.1817	-0.0437	-0.0543	-0.0207	-0.0379	-0.0592	-0.2960	-0.0841	-0.1555	-0.1555	-0.0791						
HOR	-0.0086	-0.0107	-0.0418	-0.0496	-0.0541	0.1961	-0.1224	-0.0347	-0.0643	-0.0643	-0.0327	-0.0682					
MULIT	-0.1239	-0.0449	-0.0725	0.0356	-0.1207	-0.0573	-0.0845	-0.0455	-0.1787	-0.0909	0.0724	-0.0162	0.0520				
RATE	-0.0257	-0.1121	-0.1957	-0.0308	0.0045	-0.0495	-0.0112	0.1105	-0.2080	0.1217	0.1970	-0.0189	-0.1079	0.0046			
SUMMER	-0.1118	-0.0589	0.2346	-0.0430	-0.0244	0.0330	0.1078	0.0160	-0.0823	0.0351	0.0351	-0.0107	0.0341	-0.0189	0.0341		
NATIONAL	0.1359	-0.0193	-0.0631	-0.0347	-0.0032	-0.0548	0.0369	-0.0793	0.0172	-0.0330	0.0014	0.1870	-0.0572	-0.0107	0.0046	-0.1865	
NEWYEAR	0.0094	0.0183	-0.0627	0.1562	-0.0192	0.0065	-0.0839	0.1036	0.1773	0.0655	-0.1618	-0.0263	0.0309	0.0341	-0.0397	-0.2684	-0.2110

The 1st week box office predication estimated model is:

$$\hat{FBO}_i = 21.4399 + \underset{(1.95)^*}{0.1984}\,M1_i - \underset{(2.06)^{**}}{0.0058}\,M2_i + \underset{(-2.16)^{**}}{0.0344}\,F1_i - \underset{(1.39)}{}$$

$$\underset{(-1.62)}{0.1714}\,F2_i - \underset{(-0.07)}{0.0038}\,DIR1_i - \underset{(-0.22)}{0.0028}\,DIR2_i -$$

$$\underset{(-0.40)}{1.5620}\,ACT_i - \underset{(-3.47)^{***}}{14.5647}\,WAR_i - \underset{(-2.19)^{**}}{9.3015}\,LOVE_i +$$

$$\underset{(0.26)}{1.5533}\,COMEDY_i - \underset{(-2.37)^{**}}{9.9987}\,DRAMA_i + \underset{(0.09)}{0.3968}\,SCI_i -$$

$$\underset{(-4.41)^{***}}{16.6225}\,HOR_i - \underset{(-2.19)^{**}}{4.2714}\,MULTI_i + \underset{(0.70)}{0.7218}\,RATE_i +$$

$$\underset{(1.99)^{**}}{7.5924}\,SUMMER_i - \underset{(-1.24)}{4.0990}\,NATIONAL_i +$$

$$\underset{(0.97)}{3.3332}\,NEWYEAR_i \qquad\qquad R^2 = 0.1697$$

The values of t for each estimator are shown in parentheses, and *, **, and *** are significant at the levels of 10%, 5%, and 1% respectively. In the total box office prediction model, 9 independent variables that were not significant at the 10% level. 10 independent variables were not significant at the 10% level in the 1st week's box office estimation model. Using the joint-test, the F values were 1.03 and 0.63. The joint null hypothesis of "coefficients are 0" cannot be rejected at a significant level of 10%.

In the total box office model, M1, WAR and HOR are significantly at 1% level. The results show that if the total box office of the last film in which the leading actor participated in is higher, the expected total box office of the film will also be higher. The coefficient of M1 indicates that there would be 0.5838 yuan more if 1 yuan more from the box office which the film's actor participated in the last film. The coefficient of WAR and HOR indicates that, when other variables remain unchanged, if the genre of the film is WAR film, the expected total box office will be 343.842 million yuan lower than that of other genres; if it is a horror film, the expected total box office will be 459.417 million yuan lower than that of other types. M2, F2, romance, RATE and SUMMER were significant at the significance level of 5%. M2 and F2 are significant, indicating that the number of films in which the actor and actress participated in before will affect the expected total box office. For every additional film in which the actor participated in, the expected total box office of the film will increase by 1.46 million yuan. However, while for every additional film in which the actress participated in, the expected total box office of the film will decrease by 4.891 million yuan. The romance coefficient is −9.3015, indicating that the expected box office of the romance genre will be 93.015 million yuan lower than that of other genres. The RATE coefficient is significant, indicating that the film critics ratings will also affect the expected box office of the film. Each1 higher score on Douban will positively affect the expected 54.99 million yuan more of total box office income of the film. The coefficient of the SUMMER variable is 26.9775, indicating that if the film is released in the summer vacation, the total box office of the film released in other seasons will increase by about 270 million yuan. The NEWYEAR coefficient is 13.0083, which is significantly lower than the significance level of 10%, indicating that if the film is released in the NEWYEAR period, it will earn about 130 million yuan more than the expected total box office of the film in other periods.

In the 1st-week box office model, WAR and HOR are significant under the significance level of 1%. If the story of movie is war, it is expected to reduce 145.647 million yuan in the 1st week than other genres. While if it is a horror movie, it will be 166.225 million yuan less in the 1st week box office. M1, M2, romance, DRAMA, MULTI and SUMMER are significant at 5% level. The factor of actor will affect the film's 1st weekend box office revenue from both horizontal (M1) and vertical (M2) aspects. The actor who participated in a higher box office of last film, will lead 0.1984 times higher 1st-week box office in the next film. The film 1st-week box office will be reduce 58000 yuan if the actor participated in each movie before. The romance and DRAMA coefficients show that if the main type of the film is romance or DRAMA, the box office is expected to decrease by about 93.015 million yuan and 99.987 million yuan respectively in the 1st week, while other variables remain unchanged. MULTI showed that for one more the genre in the movie, the box office will decline by 42.714 million yuan in the 1st week. The coefficient of the SUMMER is 7.5924, indicating that if the film is released in the summer vacation, the expected total box office of the film will more 75.924 million yuan than films released in other schedule.

The empirical results of this paper also have many shortcomings. For example, directors, comedy and other factors that can significantly influence the box office in general empirical judgment cannot be significantly at the significance level of 10% in this study. It is possible that the research samplehas defects itself. The data collection of China's film industry is very difficult. Many relevant data of films cannot be collected through public information, and the budget of each film is not standardized and unified. However, with the increasing supervision of the film industry, relevant data will be more open and standardized, and this problem will be effectively solved. The R-squares of both empirical models are not high, which are 0.1795 and 0.1697 respectively. The reason should be that many other factors influencing the box office are not included in the model. Hence, lack of other important factors would limited the explanatory power of the models. There is a lot of factors influencing the box office of films, and many of the factors are difficult to quantify. It is hardy to give a full analysis and a comprehensively explanation up to now. However, with the big data as well as artificial intelligence technology advances, more box-office impact factors would be found and added to the box office prediction model. The movie box office can be predicted more accurately.

4.3. Prediction model prediction results

This paper uses the box-office prediction model to make box office predictions for the four films released in 2018: "Till the end of the world," "Animal World," "Hidden man" and "The island." The prediction results are shown in Table 3.

"Till the end of the world" was released on February 1, 2018. The type of the story is Romance. Douban critics rating score is 6.7 points; the male starring is Zhao Youting who had a total of 9 films participated in before. The latest movie that he participated in was "the Warrior's Gate" which total box office is 2.283 million yuan; the female starring is Yang Zishan, and she had participated 12 film works in the past. Substituting the corresponding values

Table 3: Prediction models results comparison table

	Actors	Actresses	Released schedule	Genre (Number of Genre)	Critics rating	Prediction box office	Actual box office	Prediction deviation%
《Till the end of the world》	《The Warriors Gate》 Zhao Youting	Yang Zishan	New Year	Romance (1)	6.7			
First-week box office	M1=2.283		NEWYEAR=1	romance=1(MULTI=1)	RATE=6.7	8.2677	8.144	1.52
Total box office	M2=9	F2=12	NEWYEAR=1	romance=1		22.7972	23.448	-2.78
《Animal World》	《Guilty of Mind》 Li Yifeng	Zhou Dongyu	Summer Vacation	Act (2)	7.3			
First-Week Box Office	M1=30.397		SUMMER=1	ACT=1(MULTI=2)	RATE=7.3	26.4854	25.302	4.68
Total Box Office	M2=7	F2=12	SUMMER=1	ACT=1		73.6378	50.891	44.70
《Hidden Man》	《WuKong》 Peng Yuyan	Zhou Yun	Summer Vacation	Act(3)	7.2			
First-Week Box Office	M1=69.654		SUMMER=1	ACT=1(MULTI=3)	RATE=7.2	29.8865	31.537	-5.23
Total Box Office	M2=26	F2=7	SUMMER=1	ACT=1		104.1235	58.348	78.45
《The Island》	《Battle of Memories》 Hang Bo	Su Qi	Summer Vacation	Comedy (1)	7.2			
First-Week Box Office	M1=29.292		SUMMER=1	COMEDY=1(MULTI=1)	RATE=7.2	30.3346	53.443	-43.24
Total Box Office	M2=41	F2=69	SUMMER=1	COMEDY=1		50.4561	135.132	-62.66
	(Hang Bo+Wang Baoqiang)/2 Su Qi		Summer Vacation	Comedy (1)	7.2			
First-Week Box Office	M1=184.530		SUMMER=1	COMEDY=1 (MULTI=1)	RATE=7.2	61.1454	53.443	14.41
Total Box Office	M2=39	F2=69	SUMMER=1	COMEDY=1		141.0549	135.132	4.38

Prediction Deviation = (Prediction box office - Actual box office)/Actual box office

of the above information from "Till the end of the world" into the estimation model, the estimated value of the 1st week's box office is 82.767 million yuan, and actual first week's box office is 81.44 million yuan. The first week's box office prediction deviation is 1.52%. The estimated total box office is 227.772 million yuan, and the actual total box office is 234.48 million yuan, and the prediction deviation is −2.78%. The two box office prediction models basically played a good role of prediction.

The bias of the first week box office prediction results of the "Animal World" and "Hidden Man" are 4.68% and -5.23%. However, The prediction deviations of total box office of the two films are 44.7% and 78.45% respectively, which is a huge gap between estimations and realities. The reason might be the released schedule is on the same released period of "Dying to Survive," which was the top word-of-mouth movie of 2018. The box office market share of "Hidden Man" and "Animal World" were crowding out by "Dying to Survive." From this phenomenon, the assumption that the competition between movies at the same released time is a great influence factor for the box office. However, this article has limited space and cannot address this issue.

The prediction results of "The Island" are thought-provoking.If only considering the chief actor, Huang Bo, whose last movie was "Battle of Memories," the value of M1 is 292.92 million yuan. The prediction deviations of the 1st-week box office and total box office are −43.24% and −62.66%. The remarkable bias make the models unconvincing. However, if Wang Baoqiang and Huang Wei, both of whom are important players in "The Island," are considered as the actor factor together, the prediction deviations of two models will drop to 14.41% and 4.38%, while the value of M1 is the mean of the box office of "Battle of Memories" and "Detective Chinatown 2" which Wang Baoqiang participated in. The empirical result can make the following hypothesis: In the box office predicting, if there are many important male actors in the movie, they should be jointly considered as the influence factor.

5. CONCLUSION

The regression model shows that if the main genre of the film is war, romance, or horror, its box office in China will be significantly inferior to other types of movies. If the type is drama, it will only affect the first week of the film's box office. In addition, multiple genres films will influence the first-week box office negatively. The factor of critics ratings of films have a significant impact on the total box office, and it made little affects on the box office in the first week. This result is very consistent with the reality, because when a movie was just released, the critics rating would not be formed in time. After a week of screening, excelllent rating could contribute to attracting more people visiting the cinema and thus increase the total box office.

The box office of the movies releasing during summer vacation is significantly higher than other released schedule in China, which is different from other countries' research results that the box offices of Christmas movies are more significant higher. The reason of the difference might be the different culture which leads the different leisure time between China and other countries. In addition, this result could be an evidence that our student group is the main force of the box office contribution. What's more, the total movie box office released during China's unique New Year stall will also be significantly higher than the movies released in other periods. Empirical results show that both male and female protagonists greatly affects on the total box office and the box office of the 1st week but with different methods.

The number of movies which the actor previous participated in has little effect on the first week's box office, although the empirical statistic results shows a significant relationship between them. The impact of the number of movies which the actress previous participated in on the 1st week's box office was not significant. The total box office of the movie which the actor last participated in is positively affecting both the total box office and the first-week box office. The number of the movies which male and female protagonists have previously participated in will positively affect the total box office of the film. The influence of the actor is positive as opposed to the effect of the leading actress. This explains why many well-known directors are more inclined to use fresh faces in the choice of the heroine.

This paper regards film as an ordinary capital asset, introduces APT into the prediction model setting of movie asset investment returns for the first time, and strengthens the theoretical basis for analyzing the risk and return of cinematographic production relationship with multivariable linear regression model, which provides a new perspective for the research in this field. Based on the empirical research of domestic and foreign movie box office prediction, this paper localizes the factors of the box office prediction model and establishes a multi-factor box office prediction model to carry out empirical research and quantitative analysis. And the empirical results and the feedback results are basically in accordance with the realistic laws. The research of box office forecasting model has great practical significance: For movie investors, the establishment of box office forecasting model, which relies on key elements such as movie project type, director, starring, and schedule before the movie is released,can allow investors to have a more intuitive judgment, greatly reducing the investment risk of investors; for the film's founding group, they can apply the box office prediction model based on the relevant factors of the movie planning to calculate the expected revenue of the project, and hence use this as a basis to finance. The box office prediction model has a key guiding significance for the film and television financial industry: All along, it has been difficult to realize the asset securitization of the film industry. The main reason is that the value of the film project and the cash flow generated in the future are difficult to measure. The box office prediction model can effectively solve this problem. So as to design film and television financial products reasonably.

REFERENCES

Chang, B.H., Ki, E.J. (2005), Devising a practical model for predicting the atrical movie success: Focusing on the experience good property. Studies in Culture and Art, 18(4), 247-269.

Dan, X. (2013), An empirical study on the box office prediction of 3D films in China. Journal of Beijing Film Academy, 4, 60-64.

Han, M. (2010), Ninety Percent of the Domestic Film Revenue Depends on Box Office and 80 Percent of the Derivatives Market is Seized by Piracy [OL]. Zhongqing Online-China Youth Daily, 2010-10-21. Daily Newspaper.

Jianling, W. (2009), Development and evolution of contemporary Western film box office prediction research. Film Art, 1, 45-49.

Jinhui, W., Siyu, Y. (2015), Research on commercial value assessment model of Chinese film copyright. Contemporary Films, 11, 73-80.

Jinhui, W., Yanfang, B., Nian, L. (2016), Empirical analysis of the box office prediction models of wally wood films. University of Journalism, 1, 9-14.

Litman, B.R., Kohl, L.S. (1989), Predicting financial success of motion pictures: The'80s experience. Journal of Media Economics, 2(2), 35-50.

Ping, S. (2005), An empirical test of arbitrage pricing theory in shenzhen stock market. Journal of Zhejiang Vocational and Technical College of Industry and Commerce, 4(3), 122-124.

Reddy, S.K., Swaminathan, V., Motley, C.M. (1998), Exploring the determinants of broadway show success. Journal of Marketing Research, 35(3), 370-383.

Ross, S.A. (1976), The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), 341-360.

Scott, S. (1994), Predicting the performance of motion pictures. Journal of Media Economics, 7(4), 1-20.

Yiyi, Y. (2013), An analysis of the factors influencing the economic success of Chinese films: New development of Chinese films from the perspective of box office prediction model. Contemporary Films, 7, 18-22.