NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Denis Y. Orlov, Evgeniy M. Ozhegov

ESTIMATING THE EFFICIENCY OF SEQUELS IN THE FILM INDUSTRY

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: ECONOMICS
WP BRP 96/EC/2015

Denis Y. Orlov[1], *Evgeniy M. Ozhegov*[2]

ESTIMATING THE EFFICIENCY OF SEQUELS IN THE FILM INDUSTRY[3]

Film industry has been under investigation from social scientists for the last 30 years. A lot of the work has been dedicated to the analysis of the sequel effect on film revenue. The current paper employs data on wide releases in the US from 2010 to 2014 and provides a new look at sequel return to the domestic box office. We apply the Heckman and nonparametric sample selection approach in order to control for the non-random nature of the sequels' sample. It was found that sequels are successful only due to the fame of the first part of the series. If the sample selection is taken into control, sequels do not excel one part movies in terms of the box office. Moreover, decomposing the main factors of sequels' overearnings compared to one part movies, we found that sequels have a less competitive environment, a higher production budget, more time being in release and the number of opened theatres.

JEL Classification: D22, L82

Keywords: sequels, sample selection, nonparametric estimation, box office

[1] National Research University – Higher School of Economics – Perm. E-mail: leaverful@gmail.com

[2] National Research University – Higher School of Economics – Perm. Senior lecturer, Department of Economics and finance; E-mail: tos600@gmail.com

[3] The article was prepared within the framework of the Academic Fund Program at the National Research University Higher School of Economics (HSE) in 2015 (grant №15-05-0063) and supported within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program

Introduction

The film production business in the US is one of the paramount industries in terms of its involvement in the cultural and economic life of the country (Basuroy et al., 2003). What is more, movies are a common illustration of an “experience product” market to which information about the quality of the product and asymmetric information between firms is highly peculiar (Eliashberg and Sawhney, 1994; Nelson, 1974).

Movie production has always been contradictory business. Breaking into the film industry seems exciting, fun, glamorous and sadly, almost impossible. On the one hand, movies are very shaky and bear lot of uncertainty about the future of any project since it is directly linked with huge investments. However, there is always another side of the story: if the movie is successful, the return on investment and the reputation that comes after does not leave all of the people involved unheeded.

A blockbuster (for instance, “Superman Returns”) can be worth more than $270 million and the production process may take up several years. In order to pay back this high investment, a movie has to be among the highest grossing films of the year and collect more the $400 million earnings. If the movie becomes successful, the producer may take a decision to provide the audience with a sequel.

Sequels are very common as they are much easier to produce (the basis is already laid down by the parent movie) and they may have a built-in audience (those who liked the parent movie) who will most likely wish to see an elongation of the story. So it may seem to be a debatable issue whether to produce a sequel or to shoot a new unique movie.

The analysis and study of the film industry and sequels is very topical both from a managerial and scientific viewpoint. Studio producers often relate to sequel films as “an idea which may be used several times in order to mitigate the risks in a product line” (Turner and Emshwiller, 1993). For instance, studios may intentionally choose the release date of a film for the summer season – the long-lasting period when studios naturally acquire about 40% of the total annual box office. In the summer of 2006, for instance, four blockbuster sequels were released, which have let the studios earn even more than the mentioned share above. Examples of those sequels are “Mission Impossible 3”, “Dead Man's Chest” (sequel to “Pirates of the Caribbean”), “The Last Stand” (from the “X-Men” franchise) and “Superman Returns”.

The same situation could be seen during the 2007 summer season, when studios set a release time for no less than ten sequels. The release list included such well-known and already appellative names as the third editions of “Pirates of the Caribbean”, “Spider-Man” and “Shrek”, as well as “Ocean's Thirteen”, “The Bourne Ultimatum”, “Rush Hour 3”, the fourth edition of “Die Hard” and the fifth edition of “Harry Potter”.

From a scientific point of view, some researchers compare sequel films to quality cues (Basuroy et al., 2006) and study the effect of whether the quality of the cue may affect the box office if performing conjointly with advertising expenditures. Other scientists compare sequels to trademark expansion of leisure products (Sood and Dreze, 2006) and demonstrate that sequel films, in contradiction to common trademark expansions, may be exposed to saturation such that “seemingly dissimilar extensions are preferred to seemingly similar extensions”.

The share of box office earnings in the total earnings of the film (including the sales of DVD and Blu-ray compact discs) is gradually increasing (because of the piracy development and different software like Netflix products (Anders, 2011)), therefore studying the factors affecting the box office is very topical.

One of those factors, as noted by many researchers, is whether a film is considered for a sequel or not. As it was found out, sequels are more likely to earn more in comparison with non-sequels. However, despite the abundance of the methods used in the past and the results obtained, none of the researchers took into account the non-random sample of movies (sequels, in particular, as their box office is highly dependent on the success of the previous part), which could have led to inconsistent estimates. Therefore, in this paper we consistently estimate the box office gap between sequels and non-sequels by controlling the sample selection of sequels.

The present paper focuses on the comparison of the gross box office sales and the sequels and non-sequels. The perspective that a sequel movie is an extension of a hedonic product is adopted, and the studio/box office data is used to address two fundamental issues.

First, we examine to what extent sequels (the extensions) are able to match or exceed the box office revenues of the non-sequels. This is an important question for studio managers who count on sequels as a risk reducing strategy in a highly competitive environment (Ravid and Basuroy, 2004). Second, this paper shows the structural difference of the gross box office sales behind the sequels and non-sequels. We answer the question whether only observed characteristics of a movie affect its box office or if there are some unobserved characteristics which let the sequels (or non-sequels) excel in terms of the gross box office sales.

Controlling for sample selection lets us expand the knowledge that has been acquired by other researchers and explore earlier unknown representation of the earnings gap between sequels and non-sequels which were. What is more this article gives a clear explanation about why sequels earn more on average (without controlling for any characteristics), extracting the explicit return to domestic box office from individual movie characteristics.

Theoretical background

There have been numerous attempts to model the box office and most papers are based on a sample collected in the USA. Basically, the box office theory initiated and started to develop very quickly after the “initial analysis of successful movies” (Smith and Smith, 1986). With a sample of movies that have earned the highest amount from rentals, researchers have tried to analyze the achievements of films based on the number of Oscars (Academy Awards) and the release date. Rentals are the net sum after the share of owners is taken away from the total box office earnings. In the UK, box office earnings data are available but the earnings distribution is not available.

This research serves as a benchmark for the further development of a movie’s success theory (researchers understand success as the commercial effect of a movie). Smith and Smith (1986) regressed the film rentals by the number of Oscars and other nominations and awards. The results found by the scientists differed from the sample data for the past three decades. Researchers interpreted this fact as the nature of changing consumer tastes. However, the fluctuation of ordinary lease squares estimates may be due to the non-normal distribution of movie earnings. They finish their research in the following way: “…it may well be possible to develop empirical models relating a film’s attributes to the likelihood of consumer demand” (Smith and Smith, 1986). This research was among the first to appear in an applied economics journal. However, there are some earlier unpublished studies related to the communications literature.

Simonet (1980) attempted to explain the performance of films in the US with reference to the commercial performance of the director’s, producer’s and stars’ previous films and the number of awards they had won. The model was estimated from a sample of rental champions and almost uniquely in the literature. It was tested by generating forecasts on fresh data. However, his forecast was not as sharp due to the shortage of appropriate factors which may potentially explain the variance in box office earnings.

Litman (1983) presents a more wide-ranging model of film revenues including genre, Motion Picture Association of America (MPAA) rating, awards and star dummy variables, as well as production cost data based on a sample of 125 films. There was no data available on films which grossed under $1 million, so Litman (1983) allocated a value of $500,000 to them. This undoubtedly has introduced a bias into his results.

Wallace et al. (1993) focused on the impact of the stars on the box office revenues of films. In order to measure the impact of a star, other factors that may affect a film’s revenue should be controlled for. The control variables they used were year of release, quality rating, parental guide rating, country of origin, length in minutes, genre and cost.

Prag and Casavant (1994) extended Smith and Smith's study both in terms of number of observations and explanatory variables employed. They argued that critical acclaim was an important signal of quality and should be included. The cost of the production may also be a signal of quality, as studios would only be willing to spend large amounts on a film that was likely to be a box office success. For estimation purposes, the final cost of producing the negative was used – This includes production costs, payments to stars, editing costs, etc. They also included the MPAA rating for each film and the genre.

Another contribution is from Sochay (1994) who introduced measures of competition between films in their opening weekend to the revenue function model. Unlike Prag and Casavant (1994), all the genre dummies were found to be insignificant, but awards and nominations and time of release (Summer or Christmas release) were found to be significant.

The impact of reviews was investigated by Hirschman and Pieros (1985). They make a distinction between reviews regarding films as an art form and the audience view of films as entertainment. They suggest that a film's aesthetic value and its entertainment value may be inversely related. There is no clear, unambiguous relationship between critical and popular acclaim. There is a question mark over the role of critics as indicators of expected utility to the prospective consumer (Cameron, 1995; Eliashberg and Shugan, 1997; Holbrook, 1999).

One more study conducted by Ravid (1999) enriched the research field by analyzing the film production industry. In his paper, Ravid mainly tests a hypothesis connected with the effect of stars in the movie on the box office. What is more, he tries to understand whether the sequel effect takes place in the expansion of the box office. In Ravid's own words, "whereas the essential attributes of most commodities can be easily described and measured, this is not the case for movies. But at each moment in time studios must select projects from among many competing proposals. The exception that proves the rule is the scramble for sequels – if a successful formula is found, it must be tried again" (Ravid 1999).

The theory of signals in the film industry was extensively examined by Basuroy et al. (2006). Among the variety of cues that the studios might use in their releasing campaign, the authors chose for their analysis two of the most widespread: the creation of sequels that use well-known trademark names (Brodesser, 2000) and advertising costs (DiOrio, 2001). The authors verified a number of propositions using simultaneous-equation modeling and a real movie database to provide a new vision on the theory of interrelationship among the exhibitors, studios and audiences. Their study took into account the endogeneity of advertising expenditures, number of theater screens and the box office earnings to investigate the formerly unknown interaction function of sequels and advertising costs on the domestic box office earnings.

The results obtained by the authors closed a research gap that was incompletely studied, considering the effect of sequels' influence (Ravid, 1999) and advertising expenditure (Elberse and Eliashberg, 2003) on box office earnings. With the exception of sequels positively affecting the first-week box office earnings, an additional thought-provoking result was about the positive interaction between advertising expenditures and sequels. This may insinuate that the quality perception is positively dependent on the same level of advertising expenditure and, as a consequence, the same level of advertising costs may lead to a larger increase of box office earnings to sequels rather than to non sequels. Therefore, studios may potentially advertise less while promoting sequel movies compared to non sequel ones.

Some other authors also verified the hypothesis and discovered that the status of the film (whether it is a sequel or not) is relevant to its success (Hennig-Thurau and Heitjans, 2004; Sood and Dreze, 2006). Hennig-Thurau et al. (2006) see a sequel movie as an element of the wider idea of "cultural resemblance" (sometimes also related to 'representation' (Hennig-Thurau et al., 2001)), which defines a film's potential to be classified into a prevailing mental group to which the consumer has an affirmative opinion. Except for sequels production, cultural resemblance can be nurtured through recreations, sketch in a form of TV series or other components of widely accepted culture like comics, computer game, novels, etc. (Simonet, 1987).

Extensive, recent reviews of sequels and their effect on audiences and revenues may be found in Sood and Drèze (2006) and Hennig-Thurau et al. (2009), so I only provide a brief review here. Some of the literature conceptualizes sequels as brand extensions and thus suggests that movie goers who liked the original would be more likely to see the sequel, thus providing an increase in first-week and total attendance. Hennig-Thurau et al. (2009) suggest that the degree of transfer depends upon how similar the sequel is to the original on such characteristics as genre and MPAA rating. Moreover, while some authors such as Sood and Drèze (2006) have focused on individual consumer reactions to such issues as satiation and variety seeking in a decision to see a sequel movie, our focus is on the broader market level effects of sequels.

Despite the fact that a lot has been done in this area and in the field of sequel efficiency estimation, some more deep insight into the problem is needed. Since none of the researchers tried to empirically test the gap between the box office sales of sequels and non-sequels it may be interesting to understand whether the sequels in general are considered to be losers (or otherwise – lucky beggars) in terms of their box office sales comparative to the analogous (in terms of characteristics) non-sequel films. Another big question is whether those gaps occur due to the effect of coefficients of individual films or due to some other factors?

Methodology

In the presence of sample selection, OLS estimation of box office equations could yield biased and inconsistent estimators (Gronau, 1974; Heckman, 1976; Heckman, 1979). It is widely recognized that the standard Heckit procedure is susceptible to identification problems and sensitivity of results to model specification and distributional assumption (Vella, 1998).

At the first stage of our analysis, we would estimate the OLS regression without focusing on the sample selection. The results obtained at this stage would become a benchmark for comparison with the results obtained on the following stages when the sample selection is taken into account. The OLS regression may be formalized as follows:

$$Y_i = X_i\beta + u_i, \tag{1}$$

where Y_i – the log of domestic box office of the *i*-th film;

X_i – vector of the i-th film's characteristics;

β – marginal effects of the films' characteristics on the box office;

u_i – independent and identically distributed errors.

The dummy variable, which reflects whether the *i*-th film is a sequel or not, may be included in order to estimate the gap in box offices. However, we would use another way to estimate the gap as we are going to decompose the predicted mean into the explained and unexplained parts using Oaxaca-Blinder decomposition (1973).

In order to take in control sequels' sample selectivity, we propose a two-step model of box office determination and propensity to make a sequel (producer's decision). The sequels' box office and propensity to shoot a sequel for movie *i* is given by:

$$d_i = 1[H_i\gamma + e_i \geq 0], \tag{2}$$

$$Y_{i,s} = X_{i,s}\beta + u_{i,s}, \;\; E\left(u_{i,s} \middle| X_{i,s}\right) = 0, \tag{3}$$

where d_i – the decision to make a sequel for an *i*-th film;

H_i – a vector of determinants of the propensity that *i*-th film will have a sequel;

$Y_{i,s}$ – is the domestic box office (in logs) of a sequel for film *i*;

$X_{i,s}$ – is a vector of determinants of the sequel box office;

γ, β – associated parameter vectors;

e_i and $u_{i,s}$ - i.i.d. error terms with joint distribution.

A usual assumption for the identification of (3) is that $X_{i,s}$ is independent from $(e_i, u_{i,s})$ and $(e_i, u_{i,s})$ has bivariate normal distribution (Heckman, 1979). Then the probability of decision to make a sequel is expressed as:

$$E(d_i = 1) = Pr(e_i \geq -H_i\gamma) = \Phi(H_i\gamma), \tag{4}$$

where $\Phi(\cdot)$ - standard normal CDF.

It is well known that the Heckman model can theoretically be identified by the nonlinearity of the Inverse Mills Ratio even if the selection equation and the main equation have identical regressors. However, it is also the case that relying solely on nonlinearity is generally viewed as taking the low (and risky) road to identification. Manski (1989; 1995) points to the inherent problems for identification in a latent variable model with exclusion restrictions such as the Heckman model. Despite these serious issues, the Heckman technique is widely used because of its simplicity.

In order to avoid any problems related to poor identification of the model, the exclusion restrictions are incorporated in the selection equation. The excluded variable should not be correlated with the dependent variable on the second step, but it has to significantly influence the decision of the producer to make a sequel. Two variables are chosen as excluded ones: 1) Dummy whether the film is based on a book: this variable is a good approximation for sequels as films are often divided into several parts if the book is rather long by itself or consists of several parts (Lord of the Rings or Harry Potter, for instance); 2) Dummy whether the film is considered to be a franchise or not. Franchise identification was taken from the BoxOfficeMojo web site.

Because d_i and e_i are related by (2) and e_i has a standard normal distribution, $E(u_{i,s}|e_i \geq -H_i\gamma)$ is simply the inverse Mills ratio $\lambda(H_i\gamma)$. Domestic box offices are observed for those films which have $d_i = 1$, so that the expected domestic box office of a film (of a sequel in this particular case) is determined according to:

$$E(Y_{i,s}|d_i = 1) = X_{i,s}\beta + E(u_{i,s}|e_i \geq -H_i\gamma) = X_i\beta + \theta\lambda_i, \quad (5)$$

$$\lambda_i = \frac{\phi(H_i\gamma)}{\Phi(H_i\gamma)}, \quad (6)$$

where λ_i – inverse Mills ratio of the previous part of the film in a series;

$\phi(\cdot)$ – standard normal density;

$\Phi(\cdot)$ – standard normal distribution function;

θ – covariance between the error terms in the equation of sequels' box office and the equation of probability of this film to be released.

However, the Heckman procedure is highly dependent on the assumption on bivariate normal distribution of the error terms in the selection and outcome equation. In order to overcome the problem with the assumption, we use a more flexible semiparametric approach which assumes arbitrary continuous joint distribution of error terms. The two-step nonparametric identification procedure was introduced by Newey (1999) and extended in Das, Newey, Vella (2003) (further it is referred to as DNV).

First, we approximate the propensity score by the power series (up to the third) of covariates by linear probability model of the producer's decision to make a sequel:

$$p = E[d = 1|H] = g_o(H), \quad (7)$$

where $g_0(H)$ is the series function of covariates that determine the producer's decision to make a sequel.

After that we estimate the outcome equation with the third degree polynomial series of additivity restriction derived as control function (as a generalization of the Heckman's lambda) for sequels' domestic box office obtained from the first step:

$$E(Y_{i,s}|d_i = 1) = X_{i,s}\beta + \vartheta\lambda_i(p), \quad (8)$$

where $\lambda_i(p)$ is the series function on probability of selection (obtained from the (7) equation). However, for the components corresponding to the probability, regularity conditions require that $0 \leq p \leq 1$, so that the estimator only uses observations where estimate of p are strictly between zero and one.

After applying the sample selection correction procedure and obtaining the results, we are interested in estimating the box office difference between sequels and non-sequels in the presence of sample selectivity. We adopt the estimated sequel structure as the nondiscriminatory, competitive norm. The parameters of (5) are separately estimated for sequels and non-sequels[4].

An application of decomposition of the mean domestic box office among sequels and non-sequels in a general way can be formalized as follows (Neuman and Oaxaca, 2004):

$$\bar{Y}_s - \bar{Y}_{ns} = \bar{X}_{ns}(\hat{\beta}_s - \hat{\beta}_{ns}) + (\bar{X}_s - \bar{X}_{ns})\hat{\beta}_s, \quad (9)$$

where $\bar{Y}$ – predicted mean log of domestic box office (among sequels (subscript *s*) and non-sequels (subscript *ns*);

$\bar{X}$ – mean vector of box office determining variables;

$\hat{\beta}$ – vector of the estimated returns to the domestic box office determinants;

$(\bar{X}_s - \bar{X}_{ns})\hat{\beta}_s$ – explained input in the difference of intergroup box office gap;

$\bar{X}_{ns}(\hat{\beta}_s - \hat{\beta}_{ns})$ – unexplained input in the difference of intergroup box office gap.

However, the (9) equation does not take into account the potential sample selection of sequels while decomposing the mean box office. As Duncan and Leigh (1980), Reimers (1983), Boymond et al. (1994) show, it can be done in the following way:

$$(\bar{Y}_s - \bar{Y}_{ns}) - (\hat{\theta}_s\bar{\lambda}_s - \hat{\theta}_{ns}\bar{\lambda}_{ns}(\cdot)) = \bar{X}_{ns}(\hat{\beta}_s - \hat{\beta}_{ns}) + (\bar{X}_s - \bar{X}_{ns})\hat{\beta}_s, \quad (13)$$

where $\hat{\theta}$ – estimate of covariance between the error terms in the selection equation determining the probability of a sequel being released and the equation of the domestic box office of sequels;

---

[4] in the non-sequel group we don't apply any sample selection correction, so the computation process reduces to one-step OLS regression

$\bar{\lambda}(\cdot)$ – estimate of the additivity restriction: either Heckman's lambda $\lambda(H_i\hat{\gamma})$ or the control function series approximation $\lambda(\hat{p})$ obtained with the help of DNV procedure[5].

Data description

The dataset covers all movies widely released[6] in the United States between January 1, 2010 and December 31, 2014[7]. For each movie released, the dataset includes the individual characteristics. The data were obtained from various sources including BoxOfficeMojo, Kinopoisk and IMDB. The sample comprises 859 movie titles with 232 sequel movies. Table 1 provides some descriptive statistics for the sample used in the analysis.

Because of the long sample period, the box office revenues and production budgets are deflated to the 2014 period to accommodate trends in the average ticket price. The average ticket prices are obtained from the BoxOfficeMojo.

The industry operates on a weekly schedule. More than 80% of the movies were released on Friday (10% on Wednesday). Much of the competition is over the weekend audience, which accounts for about 70% revenues. A typical year in the movie industry (as shown in Einav (2007)) is thought to consist of four periods: summer (roughly, Memorial Day to Labor Day), holiday (Thanksgiving to mid-January), winter/spring, and fall. The first two are generally thought of as high-demand periods and the releases of big-budget movies are concentrated around a few specific weeks of the year – Memorial Day, Forth of July, Thanksgiving, and Christmas – which fall in the beginning of the summer and in the winter holiday period. Therefore, we create weekly dummies for those most important holidays which may shift up the demand.

As shown in previous studies, stars in the movies make a significant impact on its total box office, so we collected the ratings of the first three main roles in the film using the IMDB "starmeter" (however, only the leading role was considered as the others were shown to cause no influence on the box office). What is more, the directors' ratings were obtained in order to control for them. As it can be seen from the table, the distributions of stars and directors are skewed to the right due to the fact that there is no upper limit in the rating system. In order to mitigate this issue, the logs of reciprocals were calculated.

[5] $\hat{\theta}_{ns}\hat{\lambda}_{ns}$ is equal to zero as this term does not show up in the demand equation for non-sequel movies
[6] films which reached 600 screens
[7] Box offices of those films which have been released at the end of the 2014 year were followed to the end of the release, so none of the films have been excluded from the sample

Table 1. Descriptive statistics for the variables

Continuous variables	Description	Mean	Median	Min	Max
Domestic box office (adjusted)	Total revenue from ticket receipts in the US (mln dollars)	91.9	55.6	0.6	639.9
Director rating	IMDB Starmeter rating (1st position is the best)	14844.4	6148.5	3	1571361
Star rating	IMDB Starmeter rating (1st position is the best)	4052.7	198.5	1	971506
Length	Length of the film (in minutes)	108.9	106	63	201
In release	Number of days in release	97	90	97.3	477
Budget (adjusted)	Total production costs	62.1	40	0.9	356.3
Theatres open	Number of maximum theatres opened for the film	2830.1	3003	204	4468
Won awards	Total number of all the awards of the film during the wide release	8.4	2	0	211
Metascore	Movie's rating by Metacritic	52.5	52	0	100
Competition	Number of films released during the same week	4.87	5	1	10
Film rating	IMDB's rating	6.4	6.5	1.6	9
Categorical variables	**Description**	**Mean**	**Number of obs.**		**Share, %**
Genre	Adventure	0.08	66		7.7
	Action	0.3	254		30
	Horror	0.08	68		7.9
	Drama	0.2	167		19
	Animation	0.09	77		9
MPAA rating	PG (Parental guidance suggested)	0.2	169		20
	PG-13 (Not for children under 13)	0.43	358		41.7
	R (Not for children under 17)	0.37	305		35.5
Intention	Whether the film get a sequel or not (dummy)	0.39	332		38.7
Holidays	Whether a film is released during holiday (dummy)	0.09	77		9
Book	Whether a film is based on book (dummy)	0.13	107		12.5
Franchise	Whether a film is based on franchise (dummy)	0.24	199		23.2

Sources: BoxOfficeMojo, IMDB, Kinopoisk, Metacritic

Among the other individual movie variables are the following: the genre (animation, drama, horror, action and adventure); MPAA rating, as it naturally takes away some part of the box office, restricting the potential share of the audience from watching the film; length of the film in minutes; time spent in release (in days); the total production budget in mln US dollars; the maximum number of theatres showing a film in a given week; total number of awards obtained during the movie release (further awards would not affect the box office as they were received after wide release ended); and metacritic rating.

In order to control for competition in the film industry, the variable showing the number of films released within the same week as the *i*-th film was constructed. Also, there are two dummy variables reflecting whether the film is based on a book and film is considered to be a part of a franchise, which was used as excluded variables in the selection equation.

As this research aims to analyze the difference in the box office revenue between sequels and non-sequels, the total sample is divided into two subsamples through a sequel indicator variable. Intention is a binary variable which reflects whether the *i*-th film will have a sequel or not. The 'preliminary' comparison of the total domestic box office distribution among sequels and non-sequels is presented on the Fig.1. It can be seen that almost everywhere (at any part of the box office distribution, except the lowest quantiles) sequels excel in terms of their earnings in comparison with non-sequels. However, there is a need to check further as it is only a preliminary visual analysis and does not give a full insight into the problem.

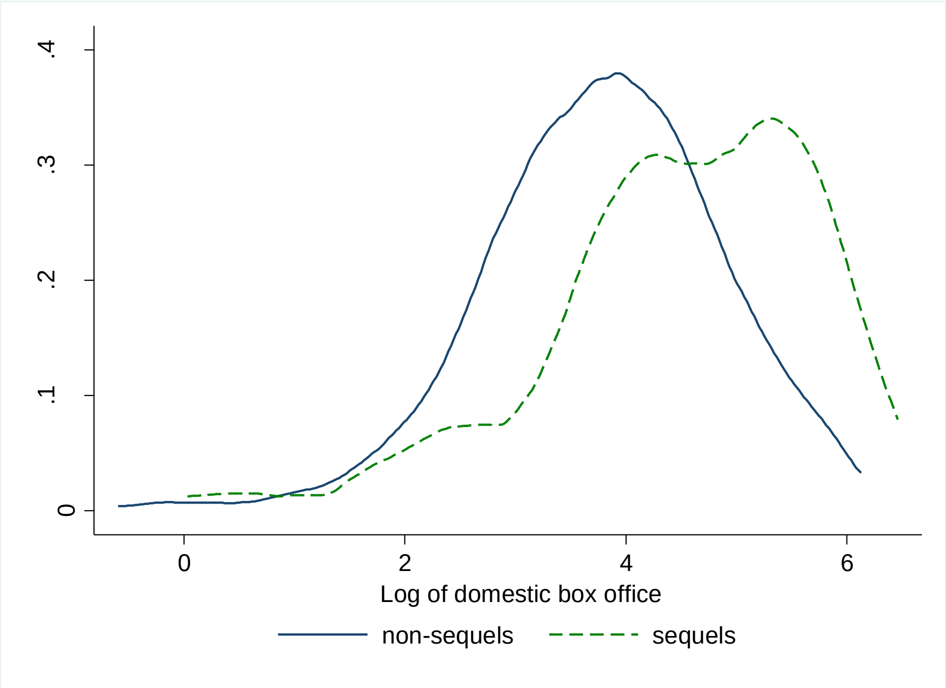

Fig. 1. Distributions of logs of total domestic box offices among sequels and non-sequels

Results

Table 2 provides the estimates of the OLS regressions of the logged domestic box office among sequels and non-sequels (the gap among the two groups is caught by the sequel dummy).

All of the obtained estimates are not contradictory regarding the signs and significance of the coefficients. Nevertheless it is rather interesting that the Holidays coefficient is insignificant in spite of its positive sign. In fact, the larger sample size may have an impact on the significance of this estimate as intuitively it seems that those films which are released during the holidays earn more.

The main coefficient of interest – the sequel dummy which depicts the gap between the sequels and non-sequels earnings – is positive and highly-significant, meaning that sequels really earn roughly 24.6% more than non-sequels in terms of box office (keeping all of the other characteristics constant). This result is consistent with all of the previous findings and suggests that people in general choose sequels just because of the 'sequelized' nature of the film: sequels are signals of high quality to an audience and people are more likely to make the movie choice in favor of a sequel.

However, the method does not take into account the important fact that a positive difference may appear not due to the fair and objective quality of sequel movies, but due to the fact that people are eager to see the continuation of the previous part and choose the movie according to this principle, increasing its earnings. If this situation is true, the positive gap in the box office among sequels and non-sequels is not a merit of an existing part of the movie but of the previous part. Therefore, we say that the sequels are getting into the sample non-randomly and propose a way of correcting the possible bias.

Table 2 also contains the results of the OLS estimation without a correction for sample selection for sequels and non-sequels separately, and the results of the estimation of the demand for sequels considering the possible sample selection via Heckman two-step procedure and the control function series approximation through the DNV procedure.

Results of the first-step estimation are not of paramount interest to the research question. The results for non-sequels are the same if the sample selection correction follows from the model setting.

As it can be seen from the uncorrected estimate of sequel and non-sequel subsamples, almost all of the significant coefficients of sequel specification are greater in absolute value than in the non-sequel specification. This fact causes the mean prediction of the sequels to be much higher, leading to a large unexplained difference in the intergroup box office difference.

The main focus here is to compare the sequel estimates obtained by the methods aimed at the elimination of sample selection with a simple OLS. As it can be seen from the results of the sample selection corrected specifications, the absolute values of estimated coefficients are pushed downwards and are much closer to the non-sequel OLS specification, meaning that the unexplained part of box office difference reduces.

Table 2. Estimates of the box office revenue equations

	Uncorrected OLS estimates			Bias corrected estimates	
				Heckit	DNV
Variable	General sample	Sequels	Non sequels	Sequels	Sequels
Action	-0.14*** (0.05)	-0.25*** (0.07)	-0.08 (0.07)	-0.24*** [0.07]	-0.26*** [0.07]
Horror	0.23** (0.11)	-0.07 (0.17)	0.33** (0.1)	-0.16 [0.16]	-0.12 (0.16)
Drama	-0.13* (0.07)	-0.36* (0.19)	-0.07 (0.07)	-0.32 [0.22]	-0.35 [0.22]
PG	-0.24*** (0.07)	-0.44*** (0.09)	-0.15** (0.07)	-0.39*** [0.09]	-0.38*** [0.09]
Competition	-0.26*** (0.03)	-0.22*** (0.05)	-0.26*** (0.04)	-0.11** [0.05]	-0.12** [0.05]
Rating	0.1*** (0.02)	0.11*** (0.04)	0.1*** (0.03)	0.08** [0.04]	0.08** [0.04]
Holidays	0.09 (0.08)	0.3** (0.16)	0.08 (0.08)	0.33* [0.18]	0.31* [0.17]
Director	0.03** (0.02)	0.07** (0.03)	0.02 (0.02)	0.06** [0.03]	0.07** [0.03]
Star	0.04*** (0.01)	0.03 (0.02)	0.04*** (0.02)	0.04* [0.02]	0.03 [0.02]
In release	1.18*** (0.06)	1.03*** (0.11)	1.19*** (0.07)	0.95*** [0.1]	0.97*** [0.11]
Theatres	0.13*** (0.02)	0.39*** (0.05)	0.12*** (0.02)	0.36*** [0.11]	0.37*** [0.11]
Budget	0.3*** (0.03)	0.29*** (0.07)	0.29*** (0.04)	0.25*** [0.08]	0.26*** [0.08]
Sequel	0.22*** (0.05)	-	-	-	-
Constant	-3.13*** (0.35)	-3.88	-3.14*** (0.41)	-3.4*** [1.01]	-3.4*** [1.01]
λ_{heckit}	-	-	-	0.15*** [0.03]	-
λ_{DNV}	-	-	-	-	-2.29 [2.52]
λ_{DNV}^2	-	-	-	-	5.2 [4.7]
λ_{DNV}^3	-	-	-	-	-2.7 [2.5]
R_{adj}^2	0.73	0.82	0.68	0.84	0.84
Number of observations	859	232	627	232	232
Number of parameters	13	12	12	13	15

Note: Dependent variable is the log of domestic box office;
*** indicates significance at 10% level, ** at 5% level, * at 1% level;
Robust standard errors in parentheses; bootstrap standard errors based on 2000 replications in brackets.
All of the continuous variables (except rating) are taken as logs

Without the correction for sample selection the coefficients are overestimated leading to a higher unexplained intergroup box office gap. The coefficient on the IMR in the Heckman model is positive and highly significant which means that the higher probability of the producer's decision about shooting a sequel is correlated with its box office revenues. This means that the more confidence a producer has in making a sequel (given the characteristics of the present part), the higher the box office of the sequel is.

As the producer's decision is highly correlated with the individual characteristics of the film (like the domestic box office, movie genre and others), the domestic box office of a sequel would depend on those through the producer's decision. In this case, the increase in box office earnings is not prone to the sequel nature of the film but to the decision of the producer of the movie. So given the individual characteristics the producer may benefit by making a profitable decision about shooting a sequel.

Another precaution was about the inconsistency of the Heckman model estimates as the errors in the selection equation and the demand for movies equation may not be jointly normally distributed. In order to loosen this premise, the nonparametric procedure of sample selection correction was applied.

As it can be seen in Table 2, the DNV estimates are very similar to the Heckman ones and their difference is statistically insignificant which gives the approval to the Heckit correction and the joint normal distribution of errors. Parameters of the control function series approximation procedure while being insignificant separately show high joint significance, however (according to the Wald test, the coefficients on these variables are jointly 0).

What is more, the two-step Heckit method is much more efficient due to its parametric nature and lower number of parameters in both of the equations (this exerts a positive influence on the asymptotic properties of estimators) which gives it more credibility over the nonparametric estimation in our scenario.

After obtaining the sample selection corrected OLS estimates we tried to analyze the gap more deeply, i.e. to separate the box office gap between sequel and non-sequels into the explained and unexplained part using Oaxaca-Blinder decomposition. Table 3 presents the decomposed domestic box office gap results of uncorrected OLS, Heckit procedure and series approximation procedure. The explained part of the difference is also decomposed into the returns of the individual variable to the box office. This analysis is not conducted for the unexplained part (Jones, 1983).

Tab. 3. Decomposition of changes in domestic box office across groups

	OLS	Heckit	DNV
Sequels' mean log of box office prediction	4.51***	4.28***	4.44***
	(0.08)	[0.08]	[0.08]
Non-sequels' mean log of box office prediction	3.8***	3.8***	3.8***
	(0.04)	[0.04]	[0.04]
Total difference	0.71***	0.48***	0.64*
	(0.09)	[0.1]	[0.39]
Explained	0.49***	0.47***	0.48***
	(0.08)	[0.1]	[0.08]
Action	-0.01	-0.01	-0.01
	(0.01)	[0.01]	[0.01]
Horror	0.01	0.01	0.01
	(0.01)	[0.01]	[0.01]
Drama	0.02*	0.02*	0.02*
	(0.01)	[0.01]	[0.01]
PG	-0.01	0.01	-0.01
	(0.01)	[0.01]	[0.01]
Competition	0.08***	0.07***	0.07***
	(0.02)	[0.02]	[0.02]
Rating	-0.01	0.01	0.01
	(0.01)	[0.01]	[0.01]
Holidays	-0.01	0.01	0.01
	(0.01)	[0.01]	[0.01]
Director	0.01	0.01	0.01
	(0.01)	[0.01]	[0.01]
Star	0.01	0.01	0.01
	(0.01)	[0.01]	[0.01]
In release	0.11**	0.11	0.11***
	(0.04)	[0.05]	[0.05]
Theatres	0.1***	0.1***	0.1***
	(0.02)	[0.02]	[0.02]
Budget	0.19***	0.19***	0.19***
	(0.03)	[0.03]	[0.03]
Unexplained	0.22***	0.01	0.16
	(0.05)	[0.07]	[0.38]

Note: *** indicates significance at 10% level, ** at 5% level, * at 1% level;
Robust standard errors in parentheses; bootstrap standard errors based on 2000 replications in brackets.
All of the continuous variables (except rating) are taken as logs

All of the conclusions and inferences that can be made from Table 3 are consistent with the ones suggested by the results of linear regressions. As we can see from Table 3, simple OLS decomposition (without correction) predicts about 24% $((e^{0.22} - 1) * 100\%)$ ((of unexplained box office difference between sequels and non-sequels which is concordant with simple OLS regression results (with the incorporation of a sequel dummy). From this fact, we can say that sequels really earn more due to some specific features of sequels.

However, as we progress in our research and move forward by taking into account the sample selection of sequels (the producer's decision to shoot a sequel), the unexplained difference disappears leaving only the explained difference, which is peculiar to individual movie characteristics. This can be seen from the obtained results of the adjusted decompositions based on Heckit and DNV two-step procedures.

Both of the sample selection models, which take into account the omitted variable bias, show that the unexplained component is statistically insignificant from zero. This corroborates the fact that sequels in the sample earn more only due to its individual features. We somehow counterfactualized the sequel as a one part movie and proved that the difference in the earnings is insignificant, meaning that the only thing that helps sequels excel in terms of their box office is the success of the previous part which is taken into account through the modeling of the producer's decision to shoot a sequel. What is more, sequels' triumph may be explained by the rational behavior of producers: none of them would be fascinated by shooting a sequel for a film that was a box office bomb or underachiever in terms of its ticket receipt revenue. Given all of the other peculiar characteristics of an individual movie are equal, a sequel would not stand out in terms of its earnings.

Another interesting point that can be drawn from the explained part of the decomposition is the return to those individual characteristics which allow us to answer the question of why still sequels' domestic box office distribution is centered to the right compared to the distribution of non-sequels. As the explained part does not really change much depending on the chosen specification (whether it is simple OLS, Heckit or DNV procedure, the coefficients in the explained part are almost the same with the same significance) we can choose one of those explained parts and visualize the return to different characteristics which allow the sequels to make a higher box office.

In Figure 2 the explained part of the Oaxaca-Blinder decomposition of return to the different characteristics is presented. Influence of the most significant variables which define the gap is depicted. Standard error bars represent the 95% significance level.

As it can be seen, most of the difference in the domestic box office (about 21%) is explained by the production budget. This is not surprising as every subsequent part of the movie series requires more investments, especially if the cast does not change but demands more royalty. Two other factors that explain about 12% of the difference are the number of opened theatres during the release and the number of days in release. The average longevity of sequels' presence in the theatres is higher than of non-sequels leading to the higher box office. A similar explanation may be applied to the number of opened theatres. One more factor which explains about 7% of the difference is competition during the first week of the movie release. Sequels are subject to be released in a less competitive time in general, therefore gaining more earnings in comparison with non-sequels.

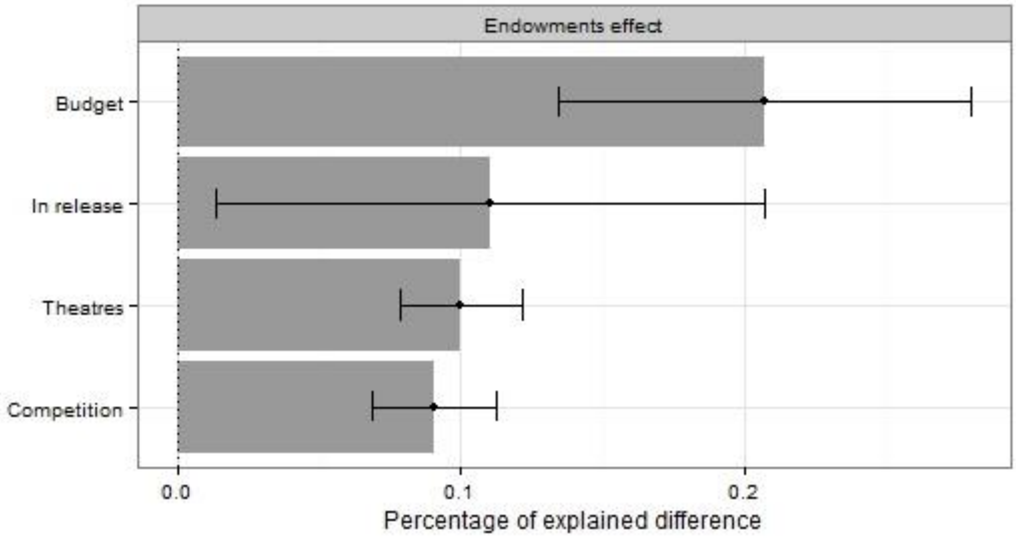

Fig. 2. Explained difference of Heckman corrected Oaxaca-Blinder decomposition

Robustness Check

If the producer's decision about shooting a sequel depends not only on the box office of the first part of the series but is known beforehand then the potential viewers know that the film will be continued despite the success of the current episode. The box office of the first part may be affected by this fact. Therefore we need to control not only the nonrandom sample selection of sequels but also the sample selectivity of non-sequels into the sample. In this scenario the endogenous decision of the producer about the following part of the movie should be taken into account while modeling the non-sequels' box office.

We are dealing with the problem by the way of Heckman correction where the selection equation models the probability of the film to become a non-sequel. The selection equation includes all of the variables considered for the sequels' selection equation, except for the domestic box office of the previous part as non-sequels do not have one. The outcome equation is the domestic box office of non-sequels regressed on the set of covariates and the selection term (which is an Inverse Mills ratio obtained from the selection equation).

Results of the outcome equation of non-sequels' Heckman correction are presented in Table 4. As it can be seen, the Heckman's lambda is insignificant meaning that the producers' decision about shooting a non-sequel is not endogenous. Comparison with the previously obtained OLS estimates lets us conclude about the invariability of estimated coefficients if the sample selection is taken into account. Therefore, we may infer that OLS estimates in this case are consistent and the most efficient. Into the bargain, we corroborate our main assumption about the producer's decision to shoot a subsequent part of the movie after obtaining the information about the box office of the first part (the decision is not taken beforehand). What is more, the results of the outcome equation show that the box office of the first part is not affected by the potential viewers' behavior, which by the supposition of the endogenous movie release may have changed. What is more, the absence of sample selectivity in the decision of non-sequel release reaffirms the main results obtained in the paper.

Table 4. Comparison of uncorrected OLS and Heckit non-sequels' estimates

Variable	Heckit	Uncorrected OLS
Action	-0.08 (0.07)	-0.08 (0.07)
Horror	0.31** (0.1)	0.33** (0.1)
Drama	-0.07 (0.07)	-0.07 (0.07)
PG	-0.15** (0.07)	-0.15** (0.07)
Competition	-0.25*** (0.04)	-0.26*** (0.04)
Rating	0.1*** (0.03)	0.1*** (0.03)
Holidays	0.08 (0.08)	0.08 (0.08)
Director	0.02 (0.02)	0.02 (0.02)
Star	0.04*** (0.02)	0.04*** (0.02)
In release	1.19*** (0.07)	1.19*** (0.07)
Theatres	0.12*** (0.02)	0.12*** (0.02)
Budget	0.28*** (0.04)	0.29*** (0.04)
Constant	-3.14*** (0.41)	-3.14*** (0.41)
λ_{heckit}	0.08 (0.16)	-
R_{adj}^2	0.68	0.68
Number of observations	627	627
Number of parameters	13	12

Note: *** indicates significance at 10% level, ** at 5% level, * at 1% level;
Robust standard errors in parentheses;
All of the continuous variables (except rating) are taken as logs

Conclusion

This paper is aimed to answer whether sequels really earn more than one part movies as it has been stated by previous findings. The research focused on the wide releases in the USA and the total domestic box office was taken as the matching benchmark. A simple comparison of domestic total box offices distributions of sequels and one part movies clearly shows that sequels are generally better off. However, an accurate comparison considers the control of different individual characteristics of the movies like budget, film or director rating, genre and others.

Previous achievements in using the conditional mean approach for comparison of box offices in this research area have shed light on the fact that sequels, after controlling for individual characteristics, earn less than is given by a simple comparison of the means. However, the absolute difference in earnings was still found to be significant in favor of sequel movies. None of the researchers, however, made a suggestion about a non-random sample of sequels, which may have been critical to the inference about the advantage of sequels in earning the higher total box office. The main aim of this paper was to check whether sequels' propensity of getting into the sample of movies is really nonrandom and if this is so, to provide new estimates concerning the interclass earnings via the correction of the non-random selection.

The main aim of this research was to control for the individual characteristic of sequels – the success of the previous part of the movie series – which is highly correlated with sequels while being unobserved, therefore, causing the estimate of sequel to be inconsistent and spurious. This issue may be interpreted as an omitted variable bias and is usually taken into consideration by the sample selection correction.

In this paper the control for the success of the previous part was introduced by the inclusion of the producer's decision about shooting a sequel into the outcome equation with the help of the Heckit procedure. For the sake of correct and consistent estimates, an additional approach which allows the error terms in the selection equation and the outcome equation to be of arbitrary joint distribution. The selection equation models the confidence the producer has about shooting another part of the series, which is highly correlated with the success of the current movie. The propensity score obtained from the selection equation is incorporated in the outcome equation as an additivity restriction in the form of an inverse Mills ratio (with the Heckit procedure) or the power series of shooting a sequel probability obtained from probability model (with nonparametric procedure).

While controlling for the main individual characteristics of the movies, the interclass gap in total domestic box office was about 20%, which is consistent with the results obtained in the previous papers. However, the inclusion of an additivity restriction into the outcome equation makes the sequel variable insignificant with highly significant and positive coefficient on lambda in the Heckit procedure. This result suggests the non-random selection of sequels into the sample and gives response to the main research question of the article about the efficiency of sequels.

In previous papers it was stated that sequels earn more just because of the fact that people are prone to visit those films because they want to see the continuation of the story. So, in fact, it was said that there is some unobserved characteristic of sequels which somehow allowed them to be discriminated by the consumers making the sequels earn more.

In this research, however, we show that it is not the consumer side but the production side which is represented through the inclusion of an omitted variable. The Heckman's lambda coefficient is highly significant and positive, which means that the expected sequel's box office is highly correlated with the probability of this sequel to be released. Because of this fact, we see prosperous sequels more frequently than sequels which perform poorly at the box office. This result corroborates the paramount idea of the current article about non-random sample of sequels. It is proved that if the most important characteristics of movies are controlled for, sequels lose any advantage holding everything else fixed in earning higher return compared to one part films.

Further analysis of the gap helped us understand why the domestic box office distribution of sequels is centered to the right in comparison with one part movies. The standard methodology which answers the question is the use of decomposition among the intergroup means. We used the Oaxaca-Blinder decomposition which allows us to divide the existing gap into the effect of characteristics (the explained part which gives an objective answer to why a sequel may earn more money) and the effects of coefficients (the unexplained part) i.e. the gap which exists despite the matter of controlling factors.

The unexplained part dissolves as we correct for sample selection of sequels and it is only the explained part of the gap that remains. Mainly, as it was found, sequels generate more box office due to the higher amounts of money invested into the production, more number of theatres involved into the release of the movie and the number of days in release.

Further research should take into account the producers' decision about the release date of the movie with the presence of week-by-week box office of individual films.

It may be mentioned, however, that any further analysis of the movie industry which is in any way aimed at dealing with sequels should be conducted with the sample selection in control, as it was shown that disregard of this fact leads researchers to the incorrect inferences.

References

Special literature

1. Ainslie, A., Drèze, X. and Zufryden, F. (2005), “Modeling movie life cycles and market share”, *Marketing Science,* Vol. 24 No. 3, pp. 508–517.
2. Basuroy, S. and Chatterjee, S. (2008), “Fast and frequent: Investigating box office revenues of motion pictures sequels”, *Journal of Business Research,* Vol. 61, pp. 798–803.
3. Basuroy, S., Chatterjee, S. and Ravid, S. (2003), “How critical are critical reviews? The box office effects of film critics, star power and budgets”, *Journal of Marketing,* Vol. 67 No. 4, pp. 103–117.
4. Basuroy, S., Desai, K. and Talukdar, D. (2006), “An empirical investigation of signaling in the motion picture industry”, *Journal of Marketing Research*, Vol. 43, pp. 287–295.
5. Blinder, A. (1973), “Wage Discrimination: Reduced Form and Structural Estimates”, *Journal of Human Resources,* Vol. 8 No. 4, pp. 436–455.
6. Blundell, R. and Powell, J. (2003), “Endogeneity in nonparametric and semiparametric regression models”, *Econometric Society Monographs,* Vol. 36, pp. 312-357.
7. Cameron, S. (1995), “On the role of critics in the culture industry”, *Journal of Cultural Economics*, Vol. 19, pp. 321-331.
8. Das, M., Newey, W. and Vella, F. (2003), “Nonparametric Estimation of Sample Selection Models,” *Review of Economic Studies*, Vol.70, pp. 33–58.
9. Duncan, G. and Leigh, D. (1980), “Wage determination in the union and nonunion sectors: A sample selectivity approach”, *Industrial and Labor Relations Review*, Vol. 34, pp. 24-34
10. Einav, L. (2007), “Seasonality in the US motion picture industry”, *The Rand journal of economics*, Vol. 38 No. 1, pp. 127-145.
11. Einav, L. (2010), “Not all rivals look alike: Estimating an equilibrium model of the release date timing game”, *Economic Inquiry*, Vol. 48 No. 2, pp. 369-390.
12. Elberse, A. and Eliashberg, J. (2003), “Demand and Supply Dynamics Behavior for Sequentially Released Products in International Markets: The Case of Motion Pictures,” *Marketing Science*, Vol. 22 No. 3, pp. 329–54.
13. Eliashberg, J. and Sawhney, S. (1994), “Modeling Goes to Hollywood: Predicting Individual Differences in Movie Enjoyment”, *Management Science*, Vol. 40, pp. 1151–73.

14. Eliashberg, J. and Shugan, S. (1997), “Film critics: influencers or predictors?” *Journal of Marketing*, Vol. 61, pp. 68-78.
15. Gronau, R. (1974), “Wage comparisons: A selectivity bias”, *Journal of Political Economy*, Vol. 82, pp. 1119-11143.
16. Heckman, J. (1974), “Shadow Prices, Market Wages and Labor Supply”, *Econometrica*, Vol. 42 No. 4, pp. 679-694.
17. Heckman, J. (1976), “The common structure of statistical models of truncation, sample selection and limited dependent variables”, *Annals of Economic and Social Measurement*, Vol. 5, pp. 475–492
18. Heckman, J. (1979), “Sample selection bias as a specification error”, *Econometrica*, Vol. 47, pp. 153–161.
19. Hennig-Thurau, T., Walsh, G. and Wruck, O. (2001), “An Investigation into the Success Factors of Motion Pictures”, *Academy of Marketing Science Review*, Vol. 5 No. 7, .
20. Hirschman, E. and Pieros, A. (1985), “Relationships among indicators of success in Broadway plays and motion pictures”, *Journal of Cultural Economics*, Vol. 9, pp. 35-63.
21. Ho, J., Dhar, T. and Weinberg, C. (2009), “Playoff payoff: Super bowl advertising for movies”, *International Journal of Research in Marketing*, Vol. 26 No. 2, pp. 168–179.
22. Holbrook, M. (1999), “Popular appeal versus expert judgments of motion pictures”, *Journal of Consumer Research*, Vol. 26, pp. 144-155.
23. Jones, F. (1983), “On Decomposing the Wage Gap: A Critical Comment on Blinder’s Method”, *Journal of Human Resources*, Vol. 18 No. 1, pp. 126-130.
24. Litman, B. (1983), “Predicting success of theatrical movies: an empirical study”, *Journal of Popular Culture*, Vol. 16, pp. 159-175.
25. Manski, C. (1989), “Anatomy of the selection problem”, *Journal of Human Resources*, Vol. 24, pp. 343-360
26. Nelson, P. (1974), “Advertising as Information”, *Journal of Political Economy*, Vol. 82 No. 4, pp. 729–54
27. Neuman, S. and Oaxaca, R. (2004), “Wage decompositions with selectivity-corrected wage equations: A methodological note”, *Journal of Economic Inequality*, Vol. 2, pp. 3-10.
28. Newey, W. (1999), “Consistency of two step sample selection estimators despite misspecifications of distribution”, *Economic Letters*, Vol. 63, pp. 129–13.
29. Newey, W. (1999), “Two-step Series Estimation of Sample Selection Models” Working paper, MIT, Department of Economics.

30. Oaxaca, R. (1973), “Male-Famale Wage Differentials in Urban Labor Markets”, *International Economic Review*, Vol. 14 No. 3, pp. 673–709.
31. Prag, J. and Casavant, J. (1994), “An Empirical Study of the Determinants of Revenues and Marketing Expenditures in the Motion Picture Industry”, *Journal of Cultural Economics*, Vol. 18, pp. 217-235.
32. Reimers, C. (1983), “Labor market discrimination against Hispanic and black men”, *The Review of Economics and Statistics*, Vol. 65, pp. 570-579.
33. Simonet, T. (1980), “Regression Analysis of Prior Experiences of Key Production Personnel as Predictors of Revenues from High-Grossing Motion Pictures in American Release”, Arno Press: New York
34. Simonet, T. (1987), “Conglomerates and Content: Remakes, Sequels, and Series in the New Hollywood, Current Research in Film: Audiences”, *Economics, and Law*, Vol. 3, pp. 154-162.
35. Smith, D. and Whan, P. (1992), “The Effects of Brand Extensions on Market Share and Advertising Efficiency,” *Journal of Marketing Research*, Vol. 29, pp. 296–313.
36. Smith, S. and Smith, V. (1986), “Successful Movies: a Preliminary Analysis”, *Applied Economics*, Vol. 18, pp. 501-507.
37. Sochay, S. (1994), “Predicting the performance of motion pictures”, *Journal of Media Economics*, Vol. 7, pp. 1-20.
38. Sood, S. and Drèze, X. (2006), “Brand extensions of experiential goods: Movie sequel evaluations”, *The Journal of Consumer Research*, Vol. 33 No. 3, pp. 352–360.
39. Turner, R. and Emshwiller, J.R. (1993), “Movie-Research Czar Is Said by Some to Sell Manipulated Findings”, *The Wall Street Journal*, December 17th, p. 1.
40. Vella, F. (1998), “Estimating model with sample selection bias: A survey”, *Journal of Human Resources*, Vol. 33, pp. 127-169
41. Vella, F. (1999), “Estimating models with sample selection bias: a survey”, *Journal of Human Resources*, Vol. 33, pp. 127–169.
42. Wallace W., Seigerman, A. and Holbrook M. (1993), “The role of actors and actresses in the success of films: how much is a movie star worth?”, *Journal of Cultural Economics*, Vol. 17, pp. 1-27.

Electronic sources:

43. Anders, C. (2011), “How much money does a movie need to make to be profitable?”, available at: http://io9.com/5747305/how-much-money-does-a-movie-need-to-make-to-be-profitable

44. BoxOfficeMojo Database, available at: http://boxofficemojo.com
45. Brodesser, C. (2000), "Sony's Twice-Told Tales", Variety, available at: https://variety.com/2000/film/news/sony-s-twice-told-tales-1117789614/
46. Charlie, J.A. (2011), "How much money does a movie need to make to be profitable?", available at: http://goo.gl/WEhpgo
47. DiOrio, C. (2001), "Marketing Is Beachhead Amid Sea of Fresh Trends", Variety, available at: https://variety.com/2001/film/news/biz-goes-to-summer-school-1117851527/
48. IMDB Movie Database, available at: http://imdb.com
49. KinoPoisk Database, available at: http://kinopoisk.ru
50. Metacritic – Movie reviews, available at: http://metacritic.com

Authors:

1. Denis Y. Orlov, National Research University Higher School of Economics (Perm, Russia).
E-mail: leaverful@gmail.com, Tel. +7 908 259 46 38
2. Evgeniy M. Ozhegov, National Research University Higher School of Economics (Perm, Russia). Senior lecturer, Department of Economics and Finance. E-mail: tos600@gmail.com, Tel. +7 952 652 45 25