# A Risk Model on the Weekend's Returns of the U.S. Movie Industry

Novriana Sumarti and Rafki Hidayat

*Abstract*— **The main purpose of a financial risk model is to make forecasts of the likely losses or gains that would be incurred for a variety of risks so it can be as references of the actual or potential investment. This paper analyzes the return of weekend's earnings of the US movie Box Office taken from January 1982 to December 2010. Based on Extreme Value Theory (EVT) analysis ([1],[2]), the data to be examined consist of only extreme values which are chosen where data surpasses a certain threshold *u*. The negative and positive tails of the data distribution will be fitted to a Generalized Pareto distribution model with particular values of parameters ξ and σ. In the next step, Value at Risk (VaR) and Expected Shortfall (ES) are computed from 2 periods, 1982 – 1995 and 1996 – 2010. How to choose an optimal threshold *u* is the key problem in GDP model. We choose *u* that makes the maximum value of VaR for each period so we can compare the results. Furthermore, for positive value of ξ, we use Goodness-of-fit test table by Zisheng and Chi [3]. Results shows that the possibility of loss for an investment in the movie industry is relatively lower than the possibility of gain for both periods of time. The values of VaR for the second period are higher than the first period. We are able to conclude that the 2008 financial crisis gave no significant effect on these measurement values in the second period. This result describes the high potential opportunity in the investment of the US movie makers.**

*Index Terms*— **Earnings Returns, Financial Risk, Extreme Value Theory, Generalized Pareto distribution, Value at Risk**

## I. Introduction

SOME reasons why the the business of motion pictures in the US attracts applied researchers in economics and commerce are the data is accessable to public and it has glamorous matters. Using statistics models, some studies have been conducted on revenues, profits, and financial returns of the film industry, and also the risk management as the measure of uncertainty in this industry. In [4] and [5], De Vany and Walls discovered that the movie industries are very risky. Half the theater screens in the United States are in bankruptcy and only 18% of movies earn a profit. Walls [6] analyzed on mobster-related movies which their profitability prospects are computed from the fitted Lévy-stable distribution. Using the stable Paretian probability model to capture the investment's uncertainty, De Vany [7] investigated that the probabilities of motion picture outcomes are much different from the Normal distribution. The tails of the distribution are "heavy" and large-scale events are much probable in a Paretian than in a Gaussian model. His paper claims that the variance is infinite and, for some variables, even the mean does not exist. Movie box office revenues, therefore, have no natural size or scale and there is no typical or average movie.

However in spite of the above gloomy picture, some movie makers took risk to make high budget projects and expected in return the movie will be categorized as blockbuster films. The U.S. weekend box office revenue is dominated by high budget movies. According to the Internet Movie Database (IMDB), among 360 blockbusters with gross box office income of over $100 million during their theatrical runs, 290 movies, or about 80%, had budgets above $60 millions . At most cases, the distribution of box office revenue is dominated by these high budget movies. However, this is not always the case. Some high budget movies may sustain losses at the box office [2].

In this paper we make a financial risk model to determine the aggregate risk in movie industries by using formal econometric techniques. The main purpose of this model is to make forecasts of the likely losses or gains that would be incurred for a variety of risks so it can be as references of the actual or potential investment. We analyse the weekend's returns of the US movie Box Office in periods of 1982-1995 and 1996-2010 using similar tools but with a different approach to [1] and [2]. The analysis is comparing the results and trying to answer a question whether the 2008 financial crisis has given an effect on the returns of the U.S box office or not.

In the next section, we will explain the definitions and techniques we used in EVT. In section 3 we will discuss how to choose an optimal threshold u in the Generalized Pareto distribution mode. In section 4 the data of the weekend's revenue of the US movie Box Office will be examined using Value at Risk (VaR) and Expected Shortfall (ES). The conclusion will be written in section 5.

## II. Extreme Value Theory

The extreme value theory (EVT) is the name for methods for modeling and measuring extreme risks. For example, it concerns the determination of the risk for the losses or gains of the weekend's returns of the US movie Box Office due to uncertainty aspects of the movie industry. The potential values of a risk have a probability distribution which we will never know exactly. We could use past losses due to similar risks, which may provide partial information about that distribution. We develop a model by selecting a particular probability distribution. We may have estimated this distribution through statistical analysis of empirical data. Simply speaking, later we could measure the risk by calculating its mean or variance. The Normal distribution is commonly used for large amount of data. However, extreme

Manuscript received March 23, 2011; revised April 14, 2011. This work was supported with a travel grant by Director General of Higher Education, Ministry of National Education Republic of Indonesia.

Novriana Sumarti is with Department of Mathematics, Institut Teknologi Bandung, Indonesia (phone: +62-22-2502545; fax: +62-22-2506420; e-mail: novriana@math.itb.ac.id).

Rafki Hidayat was with Department of Mathematics, Institut Teknologi Bandung, Indonesia (e-mail: oki_marsstar@yahoo.com).

events occur when a risk takes values from the tail of its distribution so the measure does not provide much information about the extreme risk. The tails of the Normal distribution are too thin to address the extreme loss. EVT is a tool which attempts to provide us with the best possible estimate of the tail area of the distribution.

Having estimated a distribution model, we calculate the quantile risk measure or Value at Risk (VaR) ([8],[9]). For a given probability and a time horizon, VaR assesses the worst-case loss, where the worst case is defined as the event with a 1-p probability. It can also be applied to the `best case' gain. VaR is defined as a threshold value such that the probability is in the given probability level. The value VaR is $VaR_p = F^{-1}(1-p)$, F is the cdf of the loss distribution and p is the given probability level. $VaR_p$ is the pth quantile of the distribution F. In this paper it will be estimated by function

$$VaR_p = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{n_u}p\right)^{-\xi} - 1\right), \qquad (1)$$

where $u$ is the chosen threshold, $n$ is the total number of data, and $n_u$ is the number of data surprassing the threshold $u$. Parameter $u$ is obtained from the selection of the threshold in POT, parameters ξ and σ are obtained from the application of Maximum Likelihood Estimation (MLE). Those parameters which will be explained in the next section.

The quantile risk measure does not take into consideration what the loss/gain will be if that 1-$p$ worst/best case event actually occurs. The expected shortfall or conditional tail expectation is defined as the expected size of a return that exceeds $VaR_p$ or $ES_p = E(X|X > VaR_p)$. The formula for estimated it is

$$ES_p = \frac{VaR_p + \sigma - u\xi}{1-\xi}. \qquad (2)$$

Having the values $VaR_p$ and $ES_p$ from the weekend's returns of the US movie Box Office, we can conclude the potential investment of US movie industry. The works in [2] and [3] had discussed this topic but we use quite different tools and analysis. For example, instead of logaritmic difference, we use return formulae defined in (3) which is more common tool in analyzing the financial data. This usage of return gives a significant different between positive and negative returns, especially in their Mean-Excess plots, which is not shown in the existing papers.

In the next section we discuss tools to choose an optimal threshold $u$ using statistics quantities $W^2$and $A^2$, and the graphical observation of the mean-excess function plots. With this threshold, the estimation values of of parameters ξ and σ can be found using Maximum Likelihood Estimation (MLE). It is the method of selecting values of the model parameters that maximize the GP pdf function as the likelihood function. We also use another approach to choose an optimal threshold $u$ for both positive and negative returns explained in the subsequently section.

## III. Optimal threshold determination

Data from the weekend's earning of the US movie Box Office is presented in the form of their return, the relative difference between values of the return from a week and of the previous week, and denoted as a random variable $X$.

$$X_i = \frac{R_{i+1} - R_i}{R_i}, i = 1, \dots, n. \qquad (3)$$

This formula is different from the one used in the previous work of [2],[3] which uses logarithmic differences. Theoretically, it can approximate formula (3). However in this research, a significant difference result occurs which helps the analysis. Now using the Peaks over Threshold (PoT) method, the observed data are surpassing a certain threshold u and will be the negative and positive tails of the complete data. For a certain value of $u$, we need to model the distribution of conditional excess $F_u$ which is

$$F_u(y) = P(X - u \le y \mid X > u), 0 \le y \le x_F - u,$$

where $y = x - u$ are the excesses, and $x_F$ is the right endpoint of data. This distribution will be fitted to a Generalized Pareto Distribution (GPD) model for specific values of its parameters because, according to the Pickands, Balkema and de Haan (PBH) theorem ([10],[11]), function $F_u(y)$ for large value of u is well approximated by the GPD, that is

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-1/\xi} & \text{if } \xi \ne 0 \\ 1 - e^{-y/\sigma} & \text{if } \xi = 0 \end{cases}$$

for $y \in [0, x_F - u]$ if $\xi \ge 0$ and $y \in \left[0, -\frac{\xi}{\sigma}\right]$ if $\xi < 0$, where ξ is a shape parameters and σ is a scale parameter. Note that this distribution was developed as a distribution that can model tails of a wide variety of distributions. In the case $\xi > 0$ GPD is heavy-tailed distribution, and if $\xi < 0$ GDP is short-tailed distribution. Maximum likelihood estimation (MLE) can be used to estimate the parameters of the GPD, $\xi$ and $\sigma$, which are the solution of the system

$$\frac{1}{\xi^2}\left(\sum_{i=1}^{n} ln(\sigma + \xi y_i) - n\, ln(\sigma)\right) - \left(\frac{1}{\xi} + 1\right)\left(\sum_{i=1}^{n} \frac{y_i}{\sigma + \xi y_i}\right) = 0,$$

$$\frac{1}{\sigma}(1 + \xi)\left(\sum_{i=1}^{n} \frac{y_i}{\sigma + \xi y_i}\right) - \frac{n}{\sigma} = 0,$$

for $\xi \ne 0$, and $\sigma = \frac{1}{n}\sum_{i=1}^{n} y_i$ for $\xi = 0$. The system of equations above comes from the first derivatives of the log-likelihood function for the GPD for a sample $y = \{y_1, y_2, \dots, y_n\}$ of exceess.

How to choose an optimal threshold u is a key problem that decides the fraction of data belonging to the tail, and therefore affects the results of the MLE of the parameters of the GPD function. There is a trade-off between the condition that the value of $u$ should be high enough to satisfy the PBH theorem, and the fact that the higher the threshold, the fewer observations are left for the estimation of the parameters. Two simple tools among other approaches are the goodness-of-fit test procedure using $W^2$ and $A^2$, and the graphical plot of the empirical mean excess function.

From [1], the goodness-of-fit test using Cramer-von Mises statistics $W^2$ and the Anderson-Darling stastic $A^2$ are adopted as a tool to proof a nul hypothesis saying that the random sample $X_1, X_2, \dots, X_n$ come from GPD model (4). We transform them to $Z_i = G_{\xi,\sigma}(X_i)$ by sorting the order of $X_1, X_2, \dots, X_n$, and then estimate values of $\xi$ and $\sigma$ using MLE. Quantities below are calculated using only nonzero values of $Z_i$:

$$W^2 = \sum_{i=1}^{n}\left(z_i - \frac{2i-1}{2n}\right)^2 + \frac{1}{12n},$$

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[log(z_i) + \log(1 - z_{n+1-i})].$$

Those values of $W^2$ and $A^2$ are compared to a table of critical values under a given significant level of $\alpha$, which can be seen in Table 1 below. The completed table can be

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

found in [1]. Note that the table is provided only for positive values of $\boldsymbol{\xi}$.

Table 1 Goodness-of-fit test values for GPD model
$\alpha = \Pr(W^2 \geq z)$ or $\alpha = \Pr(A^2 \geq z)$

| $\xi\backslash\alpha$ | | 0.500 | 0.25 | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 |
|---|---|---|---|---|---|---|---|---|
| 0.00 | $W^2$ | 0.057 | 0.086 | 0.124 | 0.153 | 0.183 | 0.224 | 0.255 |
| | $A^2$ | 0.397 | 0.569 | 0.796 | 0.974 | 1.158 | 1.409 | 1.603 |
| 0.10 | $W^2$ | 0.055 | 0.081 | 0.116 | 0.144 | 0.172 | 0.210 | 0.240 |
| | $A^2$ | 0.386 | 0.550 | 0.766 | 0.935 | 1.110 | 1.348 | 1.532 |
| 0.20 | $W^2$ | 0.053 | 0.078 | 0.111 | 0.137 | 0.164 | 0.200 | 0.228 |
| | $A^2$ | 0.376 | 0.534 | 0.741 | 0.903 | 1.069 | 1.296 | 1.471 |
| 0.30 | $W^2$ | 0.052 | 0.076 | 0.108 | 0.133 | 0.158 | 0.193 | 0.220 |
| | $A^2$ | 0.369 | 0.522 | 0.722 | 0.879 | 1.039 | 1.257 | 1.426 |

In the second tool, the plot of the Empirical Mean Excess function $e(u) = E(X - u|X > u)$ is calculated as in (Bi and Giles 2006, 2009):

$$e(u) = \frac{\sigma+\xi u}{1-\xi}, \text{ for } \xi < 1, \sigma + \xi u > 0. \qquad (6)$$

The mean excess function e(u) can be estimated using the following sample empirical mean excess function

$$e_n(u) = \frac{\sum_{i=1}^{n}(x_i - u)}{\sum_{i}^{n} 1_{\{x_i > u\}}} \qquad (7)$$

where $\sum_i^n 1_{\{x_i>u\}}$ is the number of data surpassing the value of u. From equation (6), $e(u)$ is a positive linear function of $u$ so is $e_n(u)$. A downward trend where $\xi < 0$ is an indication of a short-tailed distributions. On the contrary, an upward trend where $\xi > 0$ is for heavy-tailed behavior [1]. According to [3], we have to select $u$ that located at the beginning of a portion of the sample mean excess plot that is roughly linear and sloping up. Unfortunately, there are many potential values of $u$ resulted from this technique and the goodness-of-fit test. In the next section, we use additional approach by finding the maximum values of VaR and ES, and then doing the analysis to answer our questions in the beginning of this paper.

## IV. Data Analysis

Figure 1 shows the average returns of the weekend's earnings taken from www.boxofficemojo.com. There will be two periods of time where the risk model will be calculated. The first period is 1982-1995 and the second is 1996-2010. In this figure, the graph from the second period shows an extreme fluctuation which will be examined whether this phenomenon can be captured by the calculation of values of VaR or not. Especially year 2008 is commonly known as the beginning of the financial crisis which gives bad impacts to many industries. Here the linear trendline shows straight line with nearly zero slope.

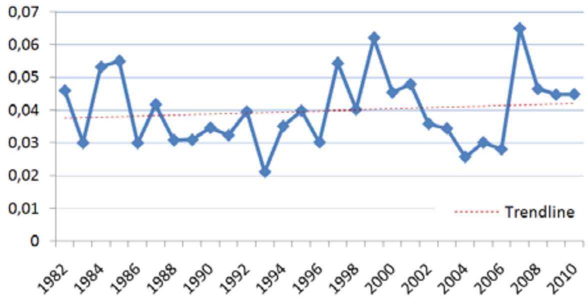


Fig. 1 Average returns of the weekends's return.

The data is divided into positive and negative returns as their distributions are asymmetric. The positive returns will depict the gain of the investment and the negative one show the loss. The observed sample comprises 1575 data, consisting 358 positive returns and 371 negative returns for period 1982-1995, and also 407 positive returns and 439 negative returns for period 1996-2010.

### A. Positive Returns

For the first period, the mean-excess of the positive returns using formula (7) is plotted in Fig. 2. The value $u$ = 0.04125 in the graph is chosen using the technique explained shortly. The graph tends to sloping up which means that formula (6) has positive value of $\xi$. It means a heavy-tailed behavior occurs for the positive returns in this period. From this fact, we will only consider values of $u$ that have corresponding positive values of $\xi$ coming from the MLE computation. There are 339 values of positive $\xi$ and denoted by index 1 to 339 based on values of $\xi$ from the smallest to the largest.

To select $\xi$ in order to have an appropiate measure of risk representing the period of time being observed, we choose the value which give the maximum value of VaR calculated from equation (1). The latter will be compared to the max value from the other period. For p = 0.01 fig. 3 shows the VaR values for all indexes. We computed them only for p = 0.01 because it distinguised significantly between those two period.

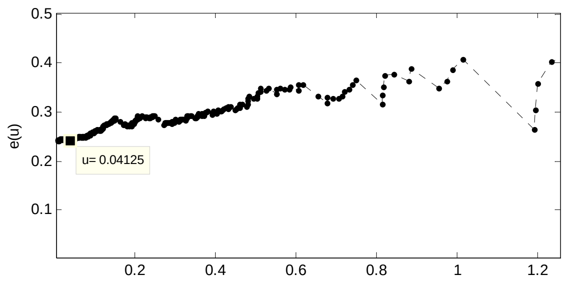


Fig.2 Mean-excess plot for positive returns 1982-1995.

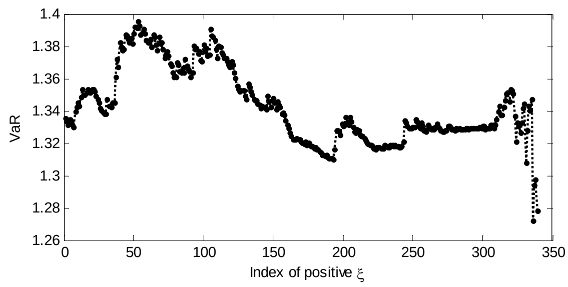


Fig. 3 The values of VaR with p=1% for positive returns 1982-1995.

The max VaR value is 1.3954, and its value of ES calculated from equation (2) is 1.9376. The corresponding value of $u$ is 0.04125 with $\xi$ = 0.1814 and $\sigma$ = 0.1982. If we calculate the goodness-of-fit test for GPD model, it has $W^2$ = 0.13099 and $A^2$ = 0.74268. Based on Table 1, the null hypothesis is accepted with $\alpha = 0.25$.

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

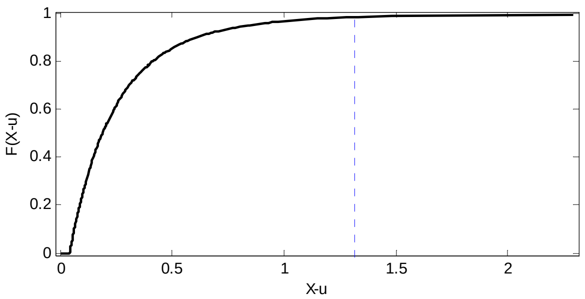


Fig.4 The *1%*-th quantile of the distribution *F* on 1982-1995 positive returns.

What do the values of VaR and ES tell us? For positive returns, it show that, with 1% probability, the returns from one weekend to the next could exceed 139.54%, and the average returns above this level will be 193.76 %. This is a very good result. In fact, there are 9 of 358 data in the first period giving return values of more than 100%, and the maximum value is 233.52 %. In an additional calculation with the same value of *u*, for probability 10%, the returns could exceed 56.01 % with the average returns of 91.71%. Figure 4 gives the description of the *1%*-th quantile of the distribution *F*.

Table 2 Numerical results

| | Positive returns | | Negative returns | |
|---|---|---|---|---|
| | 1982-1995 | 1996-2010 | 1982-1995 | 1996-2010 |
| $u$ | 0.04125 | 0.24240 | 0.26423 | 0.2755 |
| $\xi$ | 0.1814 | 0.2658 | -0.3586 | -0.3974 |
| $\sigma$ | 0.1982 | 0.1946 | 0.1374 | 0.1664 |
| VaR | 1.3954 | 1.4718 | 0.5130 | 0.5647 |
| ES | 1.9376 | 2.1819 | 0.5485 | 0.6016 |

We do the same techniques to give results for the second period, 1996-2010. All results for all clasifications are shown in Table 2. The analysis for the negative returns will be explained later. In the second period, a heavy-tailed behavior also occurs for the positive returns. The chosen value of *u* gives value 1.4718 as the max VaR of this period. It means that with 1% probability, the returns from one weekend to the next could exceed 147.18%, and the average returns above this level will be 218.19 %. These values are higher than of the previous period. Data shows 12 of 407 values of returns exeeding 100% and the max value is 300%. So it is no surprise that the analysis yields the higher values of Var and ES. Using the same value of u and a different value p = 10%, the value of VaR is 57.4% and ES is 95.9%, which are also higher than of the preceeding period of time.

### B. *Negative Returns*

Now we discuss the negative returns of both periods. The mean-excess using formula (7) for both periods are plotted. Both plots have the same behaviour. Figure 5 is for the first period. Here the graph tends to sloping down which means that GDP formula (6) has negative value of $\xi$, so values of $u$ being considered have negative values of $\xi$ resulted from the MLE computation. The negative value is an indication of a short-tailed distributions. Now to choose the optimal values of $u$, we calculated VaR values for p = 0.01. In Fig. 6, the plot of VaR values for both periods of time are presented.

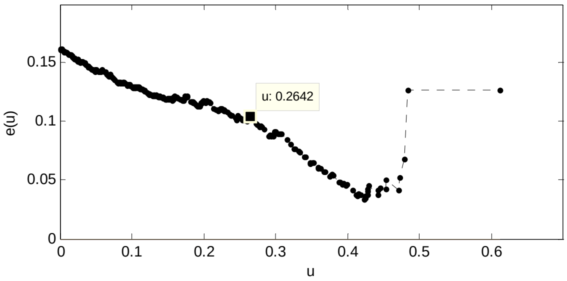


Fig. 5 Mean-excess plot for negative returns 1982-1995

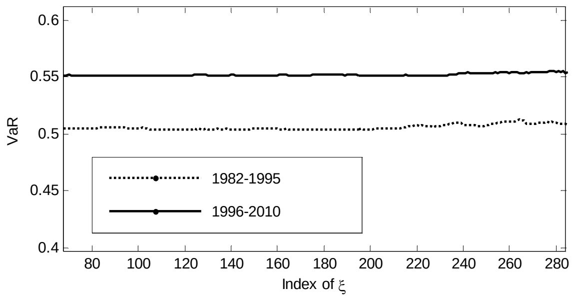


Fig.6 The values of VaR with p=1% for negative returns for 1982-1995 and 1996-2010

All values of VaR for the first period is smaller than of the second period. This also reflects on their max values which are written in Table 2. Those values means there is 1% probability that the loss from one weekend to the next could exceed 51.30% in period 1982-1995 and 56.47% in period 1996-2010. It shows higher loss in the latter period although the values are not very much different. The average loss above the corresponding VaR levels will be 54.85% (1982-1995) and 60.16% (1996-2010).

## V. CONCLUSIONS

Model of the tails of the GDP distribution for weekend movie box office returns in the US using extreme value theory shows heavy-tails for the positive returns, and short-tails for the negative returns. The positive and negative returns from period 1982-1995 have higher VaR values than of period 1996-2010, which means the possibility of gain for an investment in the movie industry in the latter period is higher than the possibility in the earlier period. However, the possibility of loss in the latter period is also higher than of in the earlier period. It can be concluded that 1998 financial crisis does not give impact on the investment in the blockbuster movie industry. In further work, we could measure the risk of the investment in the overall returns of movie industry, not only for the blockbuster movies, which is already proven a good investment in entertainment industry.

## ACKNOWLEDGMENT

We thank Dr. K.I.A. Syuhada for helping the initial work of this research. We are also grateful to Prof. D.E. Giles for email correspondences during the beginning of the research.

## REFERENCES

[1] Zisheng,O, and Chi, X., "Generalized Pareto Distribution fit to Medical Insurance Claims Data", *Appl. Math. J. Chinese Univ.* Ser. B, 2006,21(1), pp. 21-29.

[2] Bi,G. and Giles, D.E., "Modelling the financial risk associated with U.S. movie box office earnings", *Mathematics and Computers in Simulation*, Vol. 79(9) 2009, pp. 2759-2766.
[3] Bi,G. and Giles, D.E., "An application of extreme value analysis to U.S. movie box office returns", *Econometrics Working Paper* EWP0705, University of Victoria, 2007.
[4] De Vany, A. S. and W. D. Walls, "Uncertainty in the movie industry: Does star power reduce the terror of the Box Office?" *Journal of Cultural Economics*, 23(4) 1999, pp. 285-318.
[5] De Vany, A.S. & Walls, W.D., "Motion picture profit, the Stable Paretian Hypothesis, and the curse of the superstar," *Journal of Economic Dynamics and Control*, 28(6), 2004, pp. 1035-1057.
[6] Walls, W.D., "Revenues, Profitability, and Returns: Clinical analysis of the market for mobster films", *International Journal of Business and Economics*, 2004, Vol. 3, No. 2, pp. 93-106.
[7] De Vany, A.S., "Chapter 19 The Movies", *Handbook on the Economics of Art and Culture*, Vol. 1, 2006, pp. 615-665.
[8] Artzner, Ph., F. Delbaen, J.-M. Eber, and D. Heath, "Thinking coherently", *RISK,* 10, November 1997, pp. 68-71.
[9] Hardy, M.R., "An introduction to Risk Measures for Actuarial applications", *Study Note by the Casualty Actuarial Society and the Society of Actuaries*, 2006.
[10] J.I. Pickands, "Statistical inference using extreme value order statistics", *Ann. Stat.* 3, 1975, pp. 119–131.
[11] A.A. Balkema, L. de Haan, "Residual life time at great age", *Ann. Probab.* 2, 1974, pp. 792–804.

ISBN: 978-988-18210-6-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)