Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 79 (2009) 2759–2766
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Modelling the financial risk associated with U.S. movie box office earnings

Guang Bi, David E. Giles *

Department of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, BC, Canada V8W 2Y2

Received 17 April 2008; accepted 20 April 2008
Available online 26 April 2008

Abstract

In this paper we use extreme value theory to model the U.S. movie box office returns, using weekly data for the period January 1982 to September 2006. The Peak over Threshold method is used to fit the Generalized Pareto distribution to the tails of the distributions of both positive weekly returns and negative returns. Tail risk measures such as value at risk and expected shortfall are computed using maximum likelihood methods. These measures can be used as indicators for the film distributors in the preparation of movie prints, or as references for actual or potential investors in the movie industry.

Keywords: Movie revenue; Extreme values; Generalized Pareto distribution; Value at risk

1. Introduction

After being adjusted for the effects of seasonality, 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 million. In 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 sustain losses at the box office. Based on absolute loss on worldwide gross, for example, each of the top five money losers had budgets of over $100 million but lost over $90 million. When these movies were released, they dragged the weekend box office returns down. In contrast, some low budget movies are box office winners. For example, the most profitable movie based on return on investment, *The Blair Witch Project*, had a budget of only $35,000 but worldwide gross earnings of $248 million. Blockbusters and losers that appear in the tails of the distribution of the weekend box office can be taken as extreme events which can be analyzed via extreme value theory (EVT).

EVT has been applied in many areas where "disasters" occur, such as earthquakes, floods, and even terrorism attacks (e.g., [7,12,19]). Many studies have analyzed the variations in financial markets with EVT. The tail behavior of financial returns series has been discussed in [4,9,10,13–17], for example.

* Corresponding author. Tel.: +1 250 721 8540; fax: +1 250 721 6214.
E-mail address: dgiles@uvic.ca (D.E. Giles).

0378-4754/$36.00 © 2008 IMACS. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.matcom.2008.04.014

We use the Peak over Threshold (POT) method and maximum likelihood estimation to fit the Generalized Pareto distribution to the tails of the distributions for both extreme positive and extreme negative box office weekly returns. We also calculate the value at risk and expected shortfall measures of return risk.

2. Basic extreme value theory

The principal asymptotic result of EVT relates to the asymptotic distribution of “block maxima”—i.e., the maximum values of blocks, or snapshots, of data from an unknown underlying distribution. The Fisher and Tippet Theorem [8] tells us that if these maxima are suitably normalized, then they converge in distribution to one of only three forms—Gumbel, Fréchet, or Weibull. This is an extreme value analogue to conventional central limit theory. These three distributions can be encompassed by a single one—the Generalized Extreme Value (GEV) family of distributions [3, pp. 45–52]. If individual data values $\{X_1, X_2, \ldots\}$, rather than blocks, are available then it is inefficient to artificially “block” them and estimate a GEV distribution. The detailed data information can be used more efficiently by modelling the distribution, F_u, of values that are “extreme” (i.e., exceed some high threshold value, u). The conditional excess distribution function is defined as

$$F_u(y) = P(X-u) \leq y | X > u, \quad 0 \leq y \leq x_F - u, \tag{1}$$

where X is a random variable, u is a particular threshold value, $y = x - u$ are the excesses (or “exceedances”), and $x_F < \infty$ is the right endpoint of the unknown population distribution, F. So,

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)}. \tag{2}$$

As the realizations of the random variable X lie mainly between 0 and u, the estimation of F in this interval is usually quite straightforward. However, the estimation of the portion, F_u, which is of interest here, can be difficult due to the fact that the number of observations above the large enough threshold might be quite limited. The following asymptotic result is a natural generalization of the GEV result for block maxima.

Theorem [2,18]: for a large class of underlying distribution functions F the conditional excess distribution function $F_u(y)$, for u large, is well approximated by

$$F_u(y) \approx G_{\xi,\sigma}(y), \quad u \to \infty,$$

where

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma} y\right)^{\frac{-1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \mathrm{e}^{\frac{-y}{\sigma}} & \text{if } \xi = 0 \end{cases} \tag{3}$$

for $y \in [0, (x_F - u)]$ if $\xi \geq 0$, and $y \in [0, -\xi/\sigma]$ if $\xi < 0$. $G_{\xi,\sigma}$ is the so-called Generalized Pareto Distribution (GPD). Defining $x = u + y$, the GPD can be written as a function of x:

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\sigma}(x-u)\right)^{\frac{-1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \mathrm{e}^{-(x-u)/\sigma} & \text{if } \xi = 0 \end{cases} \tag{4}$$

where u is the threshold, ξ is the shape parameter and σ is the scale parameter. Maximum likelihood estimation (MLE) can be used to estimate the parameters of the GPD after selecting an appropriate threshold u. Then we can fit the GPD to the exceedances, that is, those values that exceed the threshold. The details follow.

3. The peak over threshold method

The Peak over Threshold method is used to obtain the distribution of exceedances above a certain threshold. The POT method involves the following steps: select the threshold u; fit the GPD function to the exceedances over u;