Insurance: Mathematics and Economics 77 (2017) 65–77

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions

Tom Reynkens [a,*], Roel Verbelen [b], Jan Beirlant [a,c], Katrien Antonio [b,d]

[a] LStat and LRisk, Department of Mathematics, KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium
[b] LStat and LRisk, Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[c] Department of Mathematical Statistics and Actuarial Science, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa
[d] Faculty of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received December 2016
Received in revised form August 2017
Accepted 11 August 2017
Available online 31 August 2017

Keywords:
Censoring
Composite model
Expectation–maximisation algorithm
Risk measurement
Tail modelling

ABSTRACT

In risk analysis, a global fit that appropriately captures the body and the tail of the distribution of losses is essential. Modelling the whole range of the losses using a standard distribution is usually very hard and often impossible due to the specific characteristics of the body and the tail of the loss distribution. A possible solution is to combine two distributions in a splicing model: a light-tailed distribution for the body which covers light and moderate losses, and a heavy-tailed distribution for the tail to capture large losses. We propose a splicing model with a mixed Erlang (ME) distribution for the body and a Pareto distribution for the tail. This combines the flexibility of the ME distribution with the ability of the Pareto distribution to model extreme values. We extend our splicing approach for censored and/or truncated data. Relevant examples of such data can be found in financial risk analysis. We illustrate the flexibility of this splicing model using practical examples from risk measurement.

1. Introduction

In several domains such as insurance, finance and operational risk, modelling financial losses is essential. For example, actuaries use models for claim sizes to set premiums, calculate risk measures and determine capital requirements for solvency regulations. This type of data is typically heavy-tailed and high losses can occur. A standard parametric distribution for the tail is a Pareto-type distribution, which is of key importance in extreme value theory (see e.g. McNeil, 1997). The Pareto distribution or the generalised Pareto distribution (GPD) is used to model exceedances over intermediate thresholds. However, they are not able to capture the characteristics over the whole range of the loss distribution which makes them not suitable as a global fit distribution (see e.g. McNeil (1997), Section 6.4 in Embrechts et al. (1997) and Section 6.2 in Beirlant et al. (2004)). It is often imperative to obtain a global fit for the distribution of losses, for example in a risk analysis where focus is not only on extreme events, or when setting up a reinsurance program. Instead of trying many different standard distributions, splicing two distributions (Klugman et al., 2012) is more suitable to model the complete loss distribution. In literature, a splicing model is also called a composite model. We hereby combine a light-tailed distribution for the body which covers light and moderate losses (the so-called attritional losses), and a heavy-tailed distribution for the tail to capture large losses. In the actuarial literature simple splicing models have been proposed. Beirlant et al. (2004) and Klugman et al. (2012) consider the splicing of the exponential distribution with the Pareto distribution. Other distributions for the body such as the Weibull distribution (Ciumara, 2006; Scollnik and Sun, 2012) or the log-normal distribution (Cooray and Ananda, 2005; Scollnik, 2007; Pigeon and Denuit, 2011) have also been used. Nadarajah and Bakar (2014), Bakar et al. (2015) and Calderín-Ojeda and Kwok (2016) investigate the splicing of the log-normal or Weibull distribution with various tail distributions. Lee et al. (2012) consider the splicing of a mixture of two exponentials and the GPD. The use of a mixture model in the first splicing component gives more flexibility in modelling the light and moderate losses. Fackler (2013) provides an overview of spliced distributions for loss modelling, and Brazauskas and Kleefeld (2016) illustrate the modelling performance of several spliced distributions on a real data example. Note that splicing has not only been considered in an actuarial context. Panjer (2006); Aue and Kalkbrener (2006); Peters and Shevchenko (2015) use this technique to model operational risk data.

The mixed Erlang (ME) distribution became popular in loss modelling because of several reasons (see e.g. Willmot and Woo, 2007; Lee and Lin, 2010; Willmot and Lin, 2011; Klugman et al.,

* Corresponding author.
E-mail addresses: tom.reynkens@kuleuven.be (T. Reynkens), roel.verbelen@kuleuven.be (R. Verbelen), jan.beirlant@kuleuven.be (J. Beirlant), katrien.antonio@kuleuven.be (K. Antonio).

http://dx.doi.org/10.1016/j.insmatheco.2017.08.005
0167-6687/