Applied Soft Computing 12 (2012) 2435–2441

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Claim reserving with fuzzy regression and the two ways of ANOVA

Jorge de Andrés-Sánchez *

Department of Business Management, Rovira i Virgili University, Avinguda de la Universitat 1, 43204, Reus, Spain

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form 17 May 2011
Accepted 18 March 2012
Available online 29 March 2012

Keywords:
Finance
Claims provisions
Run-off triangle
Kremer's claim reserving method
Fuzzy logic
Fuzzy numbers
Fuzzy regression

ABSTRACT

The mutant and uncertain behaviour of insurance environments does not make advisable to use a wide data-base when calculating claim reserves and so, quantifying provisions with fuzzy numbers becomes suitable. This paper firstly describes the fuzzy least squares regression that will be used in posterior developments. Subsequently we expose a claim reserving method that combines fuzzy regression with the classical statistical scheme based on two ways of ANOVA. Finally we develop a numerical application where we show in detail how to use our method to fit expected claiming costs and their variability and compare its results with those from conventional ANOVA two ways.

1. Introduction

It is not advisable to use a wide data-base to calculate claim reserves. In [24] it is pointed out that data too far from the present can lead to unrealistic estimates. For example, if claims are related to bodily injuries, the future losses for the company will depend on the growth of the wage index (which will be used to determine the amount of indemnification due), changes in court practices and public awareness of liability matters. In the actuarial field, Fuzzy Sets Theory (FST) has been used to model situations that require a great deal of actuarial subjective judgement and problems for which the information available is scarce or vague. A panoramic review about FST applications in Actuarial Science can be consulted in [22].

We think that one of the most interesting areas of FST for actuaries is Fuzzy Data Analysis (FDA). As Statistics, FST provides several techniques for searching and ordering the information contained in empirical data (e.g. for grouping elements, to find relations between variables, etc.). Within an actuarial context, FDA has been used in several areas. Horgby et al. in [10] use FDA for underwriting and reinsurance decisions, whereas [6,9,32] propose using FDA for ratemaking and risk clustering. Likewise, [1,13,14,28] adjust functions of actuarial interest with FST techniques. Specifically, [28] uses Fuzzy Mathematical Programming to fit the coefficient of dependence in the Fearlie–Gumbel–Morgenstein distribution function. Refs. [1,14] use two different fuzzy regression (FR) methods to fit the temporal structure of interest rates whereas [13] develops a FR methodology to forecast mortality with a Lee–Carter model. The present paper proposes a claim reserving method that mixes FR with the methodology developed by Kremer [15], which has been used intensively in actuarial literature (see [5,20] among others).

Determining the variability of claim provisions is as important as obtaining their fair value. This fact motivated the rise of stochastic reserving methods back in the mid 1970s. These are more complete and sophisticated than classical reserving methods (see [11] for a wide survey) because they provide not only a mean value for the reserves but also an estimate of their uncertainty (usually by means of the standard deviation). Given that the central hypothesis of these methods is that the evolution of claims is random, an ideal method must determine the value of claim provisions as a random variable completely described by its distribution function. Thus, we propose a method that calculates the expected value of reserves following [15] but, on the other hand, we estimate the variability of reserves using fuzzy numbers (FNs) instead of random variables.

In our opinion, FR has another advantage over traditional techniques. The predictions obtained after the coefficients have been adjusted are not random variables, which are difficult to use in arithmetical operations, but are FNs, which are easier to handle arithmetically. So, when starting from magnitudes estimated by random variables (e.g. which have been predicted by a statistical regression) these random variables are often reduced to their mathematical expectation so that they are easier to handle. If FNs are used, this loose of information is not needed since arithmetical

* Tel.: +34 977759832; fax: +34 977759810.
E-mail address: jorge.deandres@urv.net

1568-4946/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2012.03.033

operations with FNs are easy to compute. We think that the reasons mentioned above explain why papers such as [8,16,18,19,33–36] suggest using FR to analyse several economic problems.

The structure of the paper is as follows. In the next section we shall describe the aspects of fuzzy arithmetic and Ishibuchi and Nii's extension [12] of the FR methods described in [21,25,26]. Section 3 develops our claim reserving method. We then show a numerical application. Finally, we state the most important conclusions of the paper.

2. Fuzzy arithmetic and fuzzy regression

2.1. Some aspects of fuzzy numbers

A fuzzy number (FN) is a fuzzy subset $\tilde{a}$ defined over real numbers. It is the main instrument used in Fuzzy Set Theory (FST) for quantifying uncertain quantities. Two properties are required for a FN. The first one is that it must be a normal fuzzy set (i.e. $\sup_{\forall x \in X} \mu_{\tilde{a}}(x) = 1$). The second is that it must be convex (i.e. its α-cuts must be convex sets[1]).

For practical purposes, triangular fuzzy numbers (TFNs) are widely used FNs since they are easy to handle arithmetically and they can be interpreted intuitively. We shall symbolise a TFN $\tilde{a}$ as $\tilde{a} = (a, l_a, r_a)$ where a is the centre and l_a and r_a are the left and right spreads, respectively. For example, a subjective judgement by an actuary such as "I expect that for the next two years the claims cost inflation rate will be around 2% and deviations no greater than 1%" may be quantified in a very natural way as (0.02, 0.01, 0.01). Analytically, a TFN can be characterised by its α-cuts, a_α, as:

$$a_\alpha = [\underline{a}(\alpha), \bar{a}(\alpha)] = [a - l_a(1-\alpha), a + r_a(1-\alpha)] \tag{1}$$

In many actuarial analyses, it is often necessary to evaluate functions (e.g. the net present value of an annuity), which we shall name $y = f(x_1, x_2, \ldots, x_n)$. Then, if $x_1, x_2, \ldots, x_n$ are not crisp numbers but the FNs $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$, $f(.)$ induces the FN $\tilde{b} = f(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$ whose membership function must be obtained from Zadeh's extension principle. Unfortunately, it is often impossible to obtain a closed expression for the membership function of $\tilde{b}$, although in many cases it is possible to obtain its α-cuts, B_α, from $a_{1_\alpha}, a_{2_\alpha}, \ldots, a_{n_\alpha}$, by doing:

$$a_\alpha = f(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)_\alpha = f(a_{1_\alpha}, a_{2_\alpha}, \ldots, a_{n_\alpha}) \tag{2}$$

In actuarial mathematics, many functional relationships are continuously increasing or decreasing with respect to all the variables involved in such a way that it is easy to evaluate the α-cuts of $\tilde{b}$. Buckley and Qu in [2] demonstrate that if the function $f(\cdot)$ that induces $\tilde{b}$ is increasing with respect to the first m variables, where $m \le n$, and decreasing with respect to the last n–m variables, then b_α is:

$$\begin{aligned} b_\alpha = [\underline{b}(\alpha), \bar{b}(\alpha)] = [&f(\underline{a_1}(\alpha), \ldots, \underline{a_m}(\alpha), \overline{a_{m+}1}(\alpha), \ldots, \overline{a_n}(\alpha)), \\ &f(\overline{a_1}(\alpha), \ldots, \overline{a_m}(\alpha), \underline{a_{m+1}}(\alpha), \ldots, \underline{a_n}(\alpha))] \end{aligned} \tag{3}$$

If a FN $\tilde{b}$ is obtained from a linear combination of the TFNs $\tilde{a}_i = (a_i, l_{a_i}, r_{a_i})$, $i = 1,\ldots,n$, i.e. $\tilde{b} = \sum_{i=1}^{n} k_i \tilde{a}_i$, $k_i \in \Re$, $\tilde{b}$ will be the TFN, $\tilde{b} = (b, l_b, r_b)$, where:

$$b = \sum_{i=1}^{n} a_i k_i, \quad l_b = \sum_{\substack{i=1 \\ k_i \ge 0}}^{n} l_{a_i}|k_i| + \sum_{\substack{i=1 \\ k_i < 0}}^{n} r_{a_i}|k_i|,$$

$$r_b = \sum_{\substack{i=1 \\ k_i \ge 0}}^{n} r_{a_i}|k_i| + \sum_{\substack{i=1 \\ k_i < 0}}^{n} l_{a_i}|k_i| \tag{4}$$

It is very usual in real insurance situations to estimate magnitudes as approximate quantities, for example, by means of a sentence like "the claim provisions must be around 2000 monetary units". Clearly, FNs can be used to represent these magnitudes. However, these magnitudes also often need to be quantified with crisp values. For example, in our context, this will occur when the definitive amount of claim provisions needs to be specified in financial statements. This paper proposes using the concept of the expected value of a FN developed by Campos and González in [3], which for an FN $\tilde{a}$, we symbolise as $EV[\tilde{a}, \beta]$. This value can be obtained by introducing the decision-maker risk aversion with the parameter β, where $0 \le \beta \le 1$:

$$EV[\tilde{a}, \beta] = (1-\beta)\int_0^1 \underline{a}(\alpha)d\alpha + \beta \int_0^1 \overline{a}(\alpha)d\alpha \tag{6a}$$

Notice that the expected value of a FN is an additive measure, and so:

$$EV\left[\sum_{i=1}^{n} \tilde{a}_i, \beta\right] = \sum_{i=1}^{n} EV\left[\tilde{a}_i, \beta\right] \tag{6b}$$

2.2. Fuzzy regression model with asymmetric coefficients

In this subsection, we will describe Ishibuchi and Nii's fuzzy regression (FR) method in [12], which is an extension of the one proposed in [25–27]. As [21], [12] mixes traditional Least Squares Regression (LS) and FR method [26] but also allow a non-symmetrical structure for the data. Likewise it should be noted that both [21] and [23] show that combining Ordinary Least Squares (OLS) regression and the "pure" FR method in [25] avoids some of the drawbacks in Tanaka's traditional FR methodology.

As in conventional linear regression, we shall assume that the explained variable is a linear combination of the explanatory variables. This relationship should be obtained from a sample of n observations $\{(y_1, x_1), (y_2, x_2), \ldots, (y_j, x_j), \ldots, (y_n, x_n)\}$ where x_j is the jth observation of the explanatory variable, which is m-dimensional: $x_j = (x_{1j}, x_{2j}, \ldots, x_{ij}, \ldots, x_{mj})$. Moreover x_{ij} is the observed value for the ith variable in the jth observation which is always crisp. So, y_j is the jth observation of the explained variable, $j = 1, 2, \ldots, n$ and may either be a crisp value or a confidence interval. In both cases, it can be represented as $y_j = [\underline{y_j}, \overline{y_j}]$, where $\underline{y_j}(\overline{y_j})$ is the lower (upper) extreme of the interval y_j. In particular, we must estimate the following fuzzy linear function:

$$\tilde{Y}_j = \tilde{a}_0 + \tilde{a}_1 x_{1j} + \cdots + \tilde{a}_m x_{mj} \tag{7}$$

where $\tilde{Y}_j$ is the estimation of y_j after adjusting $\tilde{a}_0, \tilde{a}_1, \tilde{a}_2, \ldots \tilde{a}_m$.

In (7), the disturbance is not introduced as a random addend in the linear relation, but is incorporated into the coefficients $\tilde{a}_i$, $i = 0, 1, \ldots, m$, whereas in the conventional OLS regression the imprecision of the relationship between dependent and independent variables is captured by the random residual term. Dubois and Prade [7] point out that introducing fuzzy sets into regression analysis allows us to handle two sources of imprecision. Firstly, we can

[1] This second requirement means that the α-cuts must be bounded intervals in the real line.