Insurance: Mathematics and Economics 52 (2013) 477–489

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

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

Quantile credibility models

Georgios Pitselis *

University of Piraeus, Department of Statistics & Insurance Science, 80 Karaoli & Dimitriou str. T. K. 18534, Piraeus, Greece

HIGHLIGHTS

- We develop links between credibility theory and quantiles.
- Quantiles are embedded within the classical Bühlmann's (1967) credibility.
- Quantiles are embedded within Hachemeister's (1975) regression credibility model.
- Credibility estimation is based on the influence function for the *p*-quantiles.

ARTICLE INFO

Article history:
Received March 2011
Received in revised form
September 2012
Accepted 19 February 2013

Keywords:
Quantile credibility
Quantile regression credibility
Influence function

ABSTRACT

In this paper, we develop links between credibility theory and quantiles. More specifically, we show how quantiles can be embedded within the classical Bühlmann's (1967) credibility model and within Hachemeister's (1975) regression credibility model. The context of influence function is also incorporated into the above two models. For each model, credibility estimators are established and applications to real data are presented.

1. Introduction

Credibility is a premium estimation technique for a group of insurance contracts in the case where we have some claim experience for that group and a lot more experience for a larger group of contracts that are similar but not exactly the same.

Bühlmann (1967) and Bühlmann and Straub (1970) established the theoretical foundation of modern credibility theory presented as a distribution free credibility estimation. The method extended in Hachemeister's (1975) regression model, where the credibility premium depends linearly on a number of risk characteristics. Jewell (1974) has shown that credibility is exact Bayesian for a certain exponential family of distributions with natural conjugate priors.

In the actuarial literature, there are a lot of papers in credibility theory within the framework of the exponential family. The reader may be referred to Bühlmann and Gisler (2005), Sundt (1999) and Goovaerts et al. (1990).

Furthermore, Landsman and Makov (1997) extended the results on the exponential family to a richer family of distributions, the exponential dispersion family, which comprises of several distributions, some of which are heavy-tailed and as such could be of significant relevance to actuarial science. Young (1997) applied decision theory to develop a credibility formula that minimizes a loss function that is a linear combination of a squared-error term and a second derivative term. Payandeh Najafabadi (2010) approximated the Bayes estimator with respect to a general loss function and general prior distribution by a convex combination of the observation mean and mean of prior (approximate credibility formula), for a family of symmetric log-concave distributions with a location parameter.

The aim of this paper is to present the credibility (empirical Bayes) estimation of quantiles. More specifically, we incorporate quantiles into the Bühlmann's (1967) classical credibility model and into Hachemeister's (1975) regression credibility model. The quantile regression objective function is a weighted sum of absolute deviation, which gives a robust measure of location, so that the estimated coefficient vector is not sensitive to outlier observations on the dependent variable. Also, when the error term is nonnormal, quantile regression estimators may be more efficient than the least squares estimation.

In the insurance industry some legislation rules indicate that some changes over time occurred across the claim distribution. Therefore, it is essential to examine these changes at different points of the distributions. For example, Australian insurance regulations (Solvency) require that a risk margin should be established at 75% percentile of the discounted value of liabilities less than the best estimate; see Pitt (2006). Furthermore, based on quantiles

* Tel.: +30 2104142026; fax: +30 2104142340.
E-mail address: pitselis@unipi.gr.

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.insmatheco.2013.02.011

the following risk measures can be estimated: (a) the value at risk (VaR), in actuarial context known as the quantile risk measure or quantile premium principle and (b) the conditional tail expectation (CTE) known as Tail Value at Risk.

When the form of the distribution of X is not specified the natural distribution-free estimator of the p-quantile, ξ_p, is the sample p-quantile, $\hat{\xi}_p$, of ξ_p. Here we shall adopt a linear empirical Bayes empirical approach. The linear Bayes estimation introduced by Robbins (1955), discussed in depth by Hartigan (1969) and Morris (1983), and the linear empirical Bayes estimation of quantiles discussed by Maritz (1989).

For the regression case, regression curve provides a summary for the average of the distribution, while the quantile regression compares several different regression curves corresponding to the various percentage points of the distributions providing a more complete picture of the set of distributions. The book of Koenker (2005) provides an overview of regression quantiles methodologies, as well as a variety of applications from economics, biology, ecology, and finance.

In the literature of actuarial science, there are some papers dealing with quantiles. The paper of Pitt (2006) demonstrates the importance of modeling quantiles given the growing interest of regulators and others in stochastic approaches to valuation of insurance liabilities and risk margins. Kudryavtsev (2009) used quantile regression for rate-making including safety loadings and described the advantages of the quantile regression approach. Gebizlioglu and Yagci (2008) constructed tolerance intervals for bivariate quantiles of the bivariate risk distributions. Denuit (2008) provided accurate approximations for the quantiles of the conditional expected present value of the payments to the annuity provider, given the future path of the Lee Carter time index. Pitselis (2009) applied regression quantile techniques to investigate the adequacy of the own funds a company requires to remain healthy and avoid insolvency. These techniques may also provide early warning of insurer insolvencies. Portnoy (1997) used regression quantiles for graduation of Australian life tables. The idea of quantiles embedded into credibility framework was presented for the first time by Pitselis (2007) at the IME conference in Piraeus 2007.

The paper is organized as follows. Section 2 presents a brief introduction of quantile functions. In Section 3 we briefly present Bühlmann's (1967) credibility model and the derivation of the classical credibility model within a quantile framework. In Section 4 we present a brief introduction on quantile regression. In Section 5 we present Hachemeister's (1975) regression model and the development of a quantile regression credibility model. Section 6 presents the influence function for the p-quantiles as well as for the regression quantiles and incorporated them for the credibility estimation. Applications to real data are presented in Section 7 and some concluding remarks are in Section 8.

2. Preliminaries on quantiles

In this section we provide a brief introduction on quantile functions and estimators of quantile, later needed for credibility estimation. For a given data set $X_1, \ldots, X_n$, let $F(x) = P[X \le x]$ the associated distribution function, which is continuous everywhere and differentiable. We want to estimate the quantile function $\xi_p,\ 0 \le p < 1$.

The quantile of a distribution is defined as

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \ge p\}. \tag{2.1}$$

Let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics of $X_1, \ldots, X_n$ and let $\widehat{\xi}_p$ denote the sample p-quantile. It has the fundamental property that for $-\infty < x < \infty$ and $0 \le p < 1$,

$$F(x) \ge p, \quad \text{if and only if} \quad \xi_p \le x.$$

When F is continuous, ξ_p satisfies

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \ge p\} \quad \text{if and only if}$$
$$F(\xi_p) = p, \quad 0 \le p < 1.$$

Given a sample $X_1, \ldots, X_n$ of a continuous random variable X, the empirical distribution function can be defined as

$$F_n(x) = \text{fraction of } X_1, \ldots, X_n \text{ that is } \le x.$$

The empirical quantile function can be defined as

$$\hat{\xi}_p = n\left(\frac{j}{n} - p\right)X_{(j-1)} + n\left(p - \frac{j-1}{n}\right)X_{(j)},$$
$$\text{for } \frac{j-1}{n} \le p \le \frac{j}{n} \text{ and } j = 1, \ldots, n. \tag{2.2}$$

Then the derivative of $\hat{\xi}_p$ is given by

$$\hat{\xi}'_p = n(X_{(j)} - X_{(j-1)}), \quad \text{for } \frac{j-1}{n} \le p \le \frac{j}{n} \text{ and } j = 1, \ldots, n.$$

We call $n(X_{(j)} - X_{(j-1)}),\ j = 1, \ldots, n$ the spacing of the sample. The most important fact about $\hat{\xi}'_p$ is that it is asymptotically exponentially distributed with mean ξ'_p. The sample spectral density of a stationary time series has an analogous property, see Parzen (1979).

Estimators of quantiles which may behave better in small samples from symmetric densities can be obtained by a shifted piecewise linear function

$$\hat{\xi}_p = n\left(\frac{2j+1}{n} - p\right)X_{(j)} + n\left(p - \frac{2j-1}{2n}\right)X_{(j+1)}$$
$$\text{for } \frac{2j-1}{n} \le p \le \frac{2j+1}{2n} \text{ and } j = 1, \ldots, n. \tag{2.3}$$

The $\hat{\xi}_p$ is undefined for $p < \frac{1}{2n}$ or $p > 1 - \frac{1}{2n}$. Now the derivative of $\hat{\xi}_p$ is given by

$$\hat{\xi}'_p = n(X_{(j+1)} - X_{(j)}) \quad \text{for } \frac{2j-1}{n} \le p \le \frac{2j+1}{2n}.$$

In parametric models, when F is of a known location scale type $F\left(\frac{x-\theta}{\sigma}\right)$, then $F(\xi_p) = F\left(\frac{\xi_p-\theta}{\sigma}\right) = p$, and therefore, $z_p = \left(\frac{\xi_p-\theta}{\sigma}\right)$, with $F(z_p) = p$ and $\xi_p = \sigma z_p + \theta$. Then estimate θ and σ by the maximum likelihood estimation or by optimal linear combinations of order statistics. After, use the goodness of fit test, for testing the null hypothesis for various specifications of z_p corresponding to well known probability distributions, including long tail distributions (Cauchy, Pareto, extreme value, etc.), or use a mixture of distributions (robust parametric model).

In a non-parametric context, estimate z_p by estimating the density quantile function or, through suitable plots of the sample quantile functions of transformations of the data. For more on quantile functions, see Parzen (1979, 2004).

3. Credibility models

In this section we briefly revisit Bühlmann's classical credibility model and provide quantile credibility estimation for this model.