Insurance: Mathematics and Economics 45 (2009) 296–304

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

Using quantile regression for rate-making

Andrey A. Kudryavtsev *

Department of Economics, St. Petersburg State University, St. Petersburg, Russia

ARTICLE INFO

Article history:
Received November 2007
Received in revised form
November 2008
Accepted 29 July 2009

Insurance branch category:
IB42

JEL classification:
C21

Subject category:
IM30

Keywords:
Regression models
Generalized linear models
Quantile regression
Confidence band
Rate-making
Quantile approach to the net premium rate-making

ABSTRACT

Regression models are popular tools for rate-making in the framework of heterogeneous insurance portfolios; however, the traditional regression methods have some disadvantages particularly their sensitivity to the assumptions which significantly restrict the area of their applications. This paper is devoted to an alternative approach – quantile regression. It is free of some disadvantages of the traditional models. The quality of estimators for the approach described is approximately the same as or sometimes better than that for the traditional regression methods. Moreover, the quantile regression is consistent with the idea of using the distribution quantile for rate-making. This paper provides detailed comparisons between the approaches and it gives the practical example of using the new methodology.

1. Introduction

The process of rate-making is one of the most important functions of insurers. This paper is devoted to the statistical estimation of net premium rates including safety loadings. Expense loadings which are used to cover business expenses and to provide a normal profit are not the subject of the paper.

In order to correctly estimate rates or tariffs, the so-called premium principles are used in actuarial practice—see, for example, Anderson et al. (2004), Bühlmann (1970), Kudryavtsev (2004) and Mack (1997). Those principles connect the net premium rates with the random future cash flows generated by insurance contracts. The techniques of rate-making are actually based on loss distributions or their moments, which are estimated using historical data with some appropriate adjustments for changing trends.

The standard approach involves a separate analysis of two parts of net premium: expected net premium (pure cost of risk) and safety loading. The former corresponds to the expected value of future losses. The traditional models for such estimation are based on assumptions of risk homogeneity and of the absence of outliers (catastrophic losses, for example). The problem is that those assumptions are often wrong in the actuarial practice.

The safety loadings (supporting the insurer's ability to cover losses) are determined from different considerations. They are usually estimated as a proportion to one of the moments of loss distribution. The popular approaches are the principle of expected value, the principle of variation and the principle of standard deviation—e.g. see Bühlmann (1970).

Mack (1997) suggested another approach in which the net premium rate as a whole may be estimated as a quantile of loss distribution. The quantile of the distribution of random variable Y is defined as $y_\theta = \inf\{y : F_Y(y) \geq \theta\}$ for a fixed probability θ. In the case of continuous distribution with strictly monotonous distribution function, the quantile can be defined as $y_\theta = F_Y^{-1}(\theta)$. If Y is a random variable of future losses and $1 - \theta$ is a proper probability of ruin during the forecasting period then the quantile can be thought of as the estimator of the net premium rate. This approach explains the needs of safety loading quite well and it does not separate the rate into two components.

* Tel.: +7 812 393 0059; fax: +7 812 273 2400.
E-mail address: kudr@AK1122.spb.edu.

0167-6687/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2009.07.010

In practice, insurer portfolios are usually not homogeneous. In other words, the objects insured are different from each other. Actuaries try to identify risk factors (covariates) that helping explain differences between object insured. Then, the net premium rates are estimated on the basis of conditional loss distributions (assuming the values of those covariates are known). This approach actually breaks the portfolio up into sub-portfolios being described with different conditional loss distributions. Each sub-portfolio is usually more homogeneous than the total insurance portfolio.

In practice, regression methods are traditional tools for the analysis of heterogeneous portfolios—see Anderson et al. (2004), Kudryavtsev (2004) and Mack (1997). Actuaries use both linear and non-linear regressions. The latter is often recommended for actuarial applications using generalized linear models—see de Jong and Heller (2008) and Anderson et al. (2004). The ordinary and generalized regression models based on the least squares estimators are widely used in different fields including actuarial one.

Nevertheless, the traditional regression models are based on assumptions that are often wrong for real insurance portfolios. It is the main disadvantage of those models and the basic constraint for their practical applications.

First of all, those models give estimates of conditional expectations. For rate-making, the latter are thought of as the estimates of expected net premium rates (pure costs of risks). And, hence, the traditional regression approach ignores safety loadings, which should be separately analyzed from the first principles.

Moreover, traditional regression models often ignore the specific feature of the insurance data used. In particular, the following is typical for the real insurance portfolios:

- The possibility of catastrophic losses.
- The dependence of insured objects on each other (e.g. cumulating risk).
- The information shortfall to verify the statistical significance of the model chosen.

The traditional regression models are not robust to the outliers and based on distributions with light tails. Moreover, they often require the independence of observations and a large sample size. Therefore, those models should be used with great caution and often need additional qualitative and quantitative information.

Recently, intensive research has been done that tries to solve those informational and technical problems. In particular, Rousseeuw et al. (1984) discussed the construction of robust statistical models for insurance applications in the traditional regression framework. The problem of dependence in insurance portfolios was intensively studied recently—e.g. Denuit et al. (2005) and Dhaene et al. (2006). Nevertheless, an integrated solution to the problem does not exist.

Moreover, the traditional regression approach does not usually break up an insurance portfolio into absolutely homogeneous groups of policies—there is a residual heterogeneity in each sub-portfolio separated on the basis of risk factors (covariates). As a result, the best model for future losses seems to be a mixture of distributions. This does not support the assumption of the traditional regression methods about a fixed type of distribution taken from a restricted *a priori* set, for example, only Gaussian or a member of the exponential family.

In this paper, a new approach to estimating net premium rates is offered. It is based on the idea of quantile regression proposed by Koenker and Bassett (1978), Koenker and Hallock (2001) and Koenker (2005). This approach offsets some disadvantages of the traditional regression methods. Moreover, it is consistent with the idea of making net premium rates as the quantile of conditional loss distributions.

The approach proposed is widely used for estimating different types of risks in the financial area, first of all, market, credit and operational risks. The integrated description of those risks is given among others by Crouhy et al. (2001) and Jorion (2001). The quantile regression is also a popular tool of estimating Value at Risk in the financial framework—see Engle and Manganelli (1999).

The method of quantile regression has been already used for insurance applications, but only to solve some technical problems. In particular, Portnoy (1997) suggested applying it for the graduation of mortality table rates. Pitt (2006) used it to estimate the claim termination rates for Income Protection insurance. The author of the paper offered earlier using the quantile regression techniques for the estimation of net premium rates— Abduramanov and Kudryavtsev (2007). Applying quantile regression for rate-making is also discussed by Pitselis (2007) in the context of credibility.

2. Quantile regression

2.1. Advantages of the quantile regression approach

The least squares method is the most common tool among regression techniques. It produces results under the assumptions of normally distributed (Gaussian) errors and, sometimes, of their independence. Alternative approaches (like the generalized linear models) are also based on quite strong assumptions about the type of distributions and about some characteristics of data.

Nevertheless, actuaries often meet much more complex situations in their practice than that is assumed during modelling process. It may make the applications of traditional regression methods being false and, hence, useless. For example, the following problems can occur:

1. *Inaccurate estimate of loss distribution*. The estimation errors based on specific features of the sample may ruin the assumptions of its homogeneity and/or of the distribution belonging to an *a priori* set. As a result, the estimate of loss distribution may be very different than the real one that generates a model error.
2. *Loss distributions with heavy tails*. In practice, there is often a need to give larger weights to the extreme values observed than it is assumed with the Gaussian case. For such situations, the least squares method may be inappropriate.
3. *A (small) number of outliers in the sample*. They are often thought to be "noise" which is impossible to separate from "regular" data on the basis of *a priori* information. It may make the sample distribution asymmetric. This also restricts the use of the traditional regression tools and requires statistical procedures that are more sensitive to such a sample "being littered."
4. *Dependence structure of the data*. If the structure of the sample is complex enough, it is difficult or even impossible to identify that structure. Hence, there are some problems in analyzing the structure with appropriate models (for example, using the variance–covariance matrix or copula).

In other words, the traditional regression methods are inappropriate for some practical situations. As a result, any conclusions based on those methods may be incorrect.

This motivates the usefulness of alternative approaches. One of them is the quantile regression method whose main idea is to replace the quadratic deviations by absolute ones. This approach has a number of advantages:

- It is a distribution-free method, which does not need any assumptions about a parametric distribution family and does not use any properties of it.
- It is robust to outliers.
- It does not require independence or a weak degree of dependence.