Insurance: Mathematics and Economics 43 (2008) 209–213

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

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

On the link between credibility and frequency premium

Catalina Bolancé [a], Montserrat Guillén [a], Jean Pinquet [b,*]

[a] Dept. d'Econometria, Estadistica i Economia Espanyola, Universitat de Barcelona, Diagonal 690, 08034 Barcelona, Spain

[b] Département d'Economie, Ecole Polytechnique, 91128 Palaiseau Cedex, France

ARTICLE INFO

Article history:
Received September 2007
Received in revised form
May 2008
Accepted 26 May 2008

JEL classification:
C14
C25

MSC:
IM31
IB40

Keywords:
Generalized linear models
Parametric and nonparametric links
Gamma process

ABSTRACT

This paper questions the equidistribution assumption for the random effects in a frequency risk model. Two models are presented, which use parametric and nonparametric links between the variance of the random effect and frequency risk. They are estimated on a Spanish automobile insurance portfolio, for which a decreasing link is obtained. Conclusions are drawn for credibility and bonus-malus coefficients.

1. Introduction

Standard actuarial models assume that the random effects used in the distributions of the number of claims are identically distributed. In this framework, the credibility granted to the history of the policyholder increases with the frequency premium. Credibility is the no claims discount for a claimless history and the increasing relationship between an actuarial bonus and an estimated risk level is quoted in various papers (e.g. Dionne and Vanasse (1989)).

There is however empirical evidence of a decreasing link between the variance of the random effect and the frequency risk for automobile insurance data (see Sections 3 and 4 with results obtained from a Spanish portfolio). In other words, the residual relative heterogeneity on claims number distributions is more important for low risks.

In this paper, the variance of the random effect applied to Poisson distributions is conditioned on frequency risk and hence on the rating factors. First, we retain a local estimation approach of a nonparametric link between the variance and the frequency.[1] Section 2 summarizes the main properties of kernel-based estimators in generalized linear models. This approach is used in Section 3 to estimate the nonparametric link. Second, a parametric power link is specified and estimated in Section 4 from the negative binomial model. Consequences of the credibility derivation are drawn in Section 5. Section 6 concludes and an appendix contains some mathematical details.

The main empirical finding is that the link between credibility (or no claims discount) and the frequency premium is lower when the equidistribution assumption of the random effect is relaxed than it is in the usual model. The opposite result is obtained for the increase in premium after a claim.

2. Kernel estimators in generalized linear models: The index model

Generalized linear models for a response variable Y and a vector of regressors $X (X \in \mathbb{R}^k)$ assume that

$$E(Y \mid X = x) = f(x'\beta), \quad \beta \in \mathbb{R}^k. \tag{1}$$

* Corresponding author. Tel.: +33 169333423; fax: +33 169333427.
E-mail addresses: bolance@ub.edu (C. Bolancé), mguillen@ub.edu (M. Guillén), jean.pinquet@polytechnique.edu (J. Pinquet).

[1] Local estimation techniques can also be used for prediction on time series (see Qian (2000) for applications to insurance.

0167-6687/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2008.05.015

The function f is a link between an index $x'\beta$ (i.e. a scalar product between regressors and parameters) and the expectation of the response variable. In the literature on generalized linear models, the link usually refers to the reciprocal of f, but here we retain the function which is estimated in the first place. For identifiability purposes detailed later, we suppose that the intercept is not included as a regressor in (1) and that the distribution of X is nondegenerate in $\mathbb{R}^k$. For basic generalized linear models, f is given and then an intercept must be included in the regression (see McCullagh and Nelder (1989)). For a count data model, f is usually the exponential function. If f is unknown in the specification so that an estimation is required, Eq. (1) is referred to as an index model (see Härdle et al. (1997)). In that case, there is an obvious identifiability conflict between β and f in Eq. (1). Only the line $\mathbb{R}\beta$ can be identified from the data. In other words, what is identified is the conditional expectation, assumed to be constant on affine hyperplanes of $\mathbb{R}^k$ which are orthogonal to a given vector. For a given value of β, a nonparametric estimation of $f(s)$ can be based on local weighted averages of the response variable, with weights which decrease with the distance between s and the individual values of the index.

A first estimation of index models can be obtained from a parametric specification of the distribution of Y defined conditionally on X. Let us assume that we have, in that case,

$$E(Y \mid X = x) = f_0(c + x'b); \quad c \in \mathbb{R},\ b \in \mathbb{R}^k, \tag{2}$$

with f_0 a given link function. Let $\widehat{b}$ be the maximum likelihood estimator of b. The conditional expectation defined in (1) can be estimated with a kernel estimator.

In a sampling model on (X, Y) with n observations $(x_i, y_i)_{i=1,\dots,n}$, an estimator of $E(Y \mid X)$ of the Nadaraya–Watson type is obtained from a kernel K (usually an even probability density function) and a bandwidth h in the following way:

$$\widehat{E}(Y \mid X) = \frac{\sum_{i=1}^{n} y_i K_h\left(X'\widehat{b} - x_i'\widehat{b}\right)}{\sum_{i=1}^{n} K_h\left(X'\widehat{b} - x_i'\widehat{b}\right)}, \qquad K_h(u) = \frac{K\left(\frac{u}{h}\right)}{h}. \tag{3}$$

The bandwidth is a smoothing parameter. The closer it is to zero, the more estimation is performed on a local basis. The estimation given in (3) exhibits an invariance property as it only depends on $\widehat{b}/h$.

A suitable value of h can be derived from a cross-validation method similar to that proposed in Härdle (1990) for the Nadaraya–Watson kernel estimator. It is equal to

$$\arg\min_h \ CV(\widehat{b}, h) = \sum_{i=1}^{n} (y_i - \hat{E}_{-i}(Y_i))^2, \tag{4}$$

where $\hat{E}_{-i}(Y_i)$ is the leave-one-out estimator[2] defined by

$$\hat{E}_{-i}(Y_i) = \frac{\sum_{j\neq i} y_j K_h\left(x_i'\widehat{b} - x_j'\widehat{b}\right)}{\sum_{j\neq i} K_h\left(x_i'\widehat{b} - x_j'\widehat{b}\right)} = \frac{\sum_{j\neq i} y_j K\left(\left(x_i - x_j\right)' \frac{\widehat{b}}{h}\right)}{\sum_{j\neq i} K\left(\left(x_i - x_j\right)' \frac{\widehat{b}}{h}\right)}. \tag{5}$$

This estimation of the conditional expectation $E(Y \mid X)$ is not necessarily consistent since it is derived from a wrong link function (f_0 instead of f). Two results on this issue are worth mentioning.

- On one hand, a consistent estimator of the conditional expectation can be obtained from the cross-validation criterion defined in (4) and (5). Replacing $\widehat{b}$ by b in (4) defines a function $CV(b, h)$ which can be minimized with respect to b and h. The optimal values of b and h are plugged into the expression for the estimated conditional expectation given in (3). Sufficient conditions for the consistency of the estimation are given in Härdle et al. (1993). However the minimization of the cross-validation criterion is cumbersome, since it necessitates a double sum on individuals, also, Eq. (5) shows that only b/h is identified. Hence an identifying constraint needs to be added in the estimation.
- On the other hand, the maximum likelihood estimation of a parametric model given in the first place can lead to consistent estimation of the conditional expectation under conditions which are first related to the distribution of the regressor X (see Li and Duan (1989), and the Appendix). Owing to the identification issue mentioned above, consistency means that $\widehat{b}$ converges towards a limit b_0 which belongs to the line $\mathbb{R}\beta$.

3. Kernel estimators for the variance of the random effect in a Poisson model

Let us consider cross-section data. The policyholders in the portfolio are indexed by $i = 1, \dots, p$. All the risk exposure durations are supposed equal (they are equal to one year in our empirical study). Frequency risk must be expressed for a time unit, otherwise the results on the link investigated in this paper would not be coherent with respect to period aggregation. We denote n_i as the number of claims reported by policyholder i, N_i the related random variable and x_i as the vector of regression components. If U_i is the random effect, the distribution of N_i in the Poisson model with random effects is obtained from an expectation taken with respect to the random effect U_i

$$P[N_i = n_i] = E\left[P_{\lambda_i U_i}(n_i)\right]; \quad \lambda_i = \exp(c + x_i'b);$$
$$P_\lambda(n) = \exp(-\lambda)\frac{\lambda^n}{n!}.$$

If the equidistribution assumption of the random effects is relaxed, we can link their variance and frequency risk and write for instance

$$E(U_i) = 1; \qquad V(U_i) = \sigma^2(\lambda_i), \qquad \lambda_i = E(N_i). \tag{6}$$

The function σ^2 must have nonnegative values. We will investigate a power link in Section 4 but we first let the data speak from a nonparametric estimation of σ^2. The starting point is the usual moment-based estimator

$$V(U_i) = \frac{V(N_i) - E(N_i)}{E^2(N_i)} = \frac{E(N_i^2) - E^2(N_i) - E(N_i)}{E^2(N_i)}. \tag{7}$$

The nonparametric link is then obtained with an index model strategy described in Section 2. First, an estimation is performed from the maximum likelihood estimation of the Poisson model with regression components

$$N_i \sim P(\lambda_i), \quad \lambda_i = \exp(c + x_i'b).$$

For each policyholder i, we obtain the index $s_i = \widehat{c} + x_i'\widehat{b}$ and the parametric frequency premium $\widehat{E}^0(N_i) = \exp(s_i)$. Then the variance of the random effect is estimated from Eq. (7) and from nonparametric estimators $\widehat{E}(N_i^m)$, $m = 1, 2$. These estimators are derived from Eq. (3), with $Y = N^m$, $m = 1, 2$. In what follows, we retain a Gaussian kernel. Hence K_h is the density of a $N(0, h^2)$ distribution. The bandwidth h_m retained for the estimation of $E(N_i^m)$ $(m = 1, 2)$ is obtained with the leave-one-out approach

[2] The individual for which the non parametric expectation is derived must be withdrawn from the computation, otherwise the bandwidth would converge towards zero in the minimization of the cross-validation criterion.