ROBUST INFERENCE IN RATING MODELS

Gilles DUPIN [1], Alain MONFORT [2] and Jean-Pierre VERLE [3]

ASTIN topic : Risk evalution

Abstract.

Various sources of inconsistency are identified in usual statistical rating models. Several semiparametric methods, which are more robust with respect to specification errors, are proposed. In particular, the Pseudo Maximum Likelihood Methods, the Generalized Method of Moments and the Asymtotic Least Squares Methods are used in a new approach of a priori and a posteriori rating. An empirical implementation, based on data from Groupe Monceau, is discussed.

Keywords :

Inconsistency, Robustness, Misspecification, Semiparametric Methods, Pseudo Maximum Likelihood Methods, Generalized Method of Moments, Asymptotic Least Squares Method, A Priori and A Posteriori rating.

[1] General Manager, Monceau-Assurances, Paris France

[2] Conservatoire National des Arts et Metiers, and CREST-INSEE, France Address : Alain MONFORT - CREST - 15 Bd Gabriel Péri - 92245 Malakoff France - Phone : +33141177728 - Fax : 33141177666 - Email : monfort@ensae.fr

[3] Consultant actuary, Société Européene d'Actuariat Dommages, Paris

1. INTRODUCTION

The standard statistical rating models are particular cases of conditional parametric models. More precisely, in these models we consider, for each observation i, a vector of endogenous variables Y_i (typically the number and the size of the claims for which policyholder $i, i = 1, \ldots, n$, is responsible during one or several time periods) and a vector of exogenous variables X_i; then the true conditional probability density function of Y_i given the observed value x_i of $X_i, f_o(y_i/x_i)$, is assumed to belong to a given parametric family $\{f(y_i/x_i;\theta), \theta \in \Theta\}$ and inference methods, usually based on the Maximum Likelihood (ML) theory, are implemented in order to derive estimators of the quantities of interest, typically the conditional moments of $g(Y_i)$ given $X_i = x_i$, where $g(.)$ is some given function.

The aim of this paper is threefold. First we identify sources of potential inconsistency in this standard approach. Indeed the estimator of the quantity of interest may be inconsistent when n goes to infinity, as soon as the model is misspecified, that is to say when $f_o(y_i/x_i)$ does not belong to the family $\{f(y_i/x_i;\theta), \theta \in \Theta\}$. We stress several reasons for which such a misspecification may occur, in particular error on the shape of the p.d.f. retained for subvectors of Y_i, error in homoscedasticity assumptions for these p.d;f. or error in independence assumptions made about some components of Y_i and, therefore, errors on the joint p.d.f. of Y_i.

The second objective of the present paper is to propose methods aiming at reducing this inconsistency risk. The recent econometric literature proposes three routes for this purpose. The first one is the nonparametric approach. In rating models it would consist in estimating nonparametrically the conditional moments of $g(Y_i)$ given $X_i = x_i$; however this method cannot be used in practice because of the large dimension of X_i which, moreover, always has qualitative components ruling out the kernel methods. Another approach consists in staying in a parametric framework and in considerably increasing the size of the parameter θ; these methods, sometimes called semi-nonparametric, have been used for modeling frequency components [see Gourieroux-Monfort (1997)] but their generalisation to both frequency and severity components is not obvious. The third approach, which will be followed here, is the semiparametric approach, in which we do not make assumptions about the conditional p.d.f. of Y_i and, instead, we specified directly the conditional moments of interest. In such a framework, the likelihood function no longer exists and we have to propose other methods. Here we will consider the Pseudo Maximum Likelihood Methods White 1981, and Gourieroux, Monfort, Trognon 1984 a), the Generalized Method of Moments (Hansen, 1982) and the Asymptotic Least Square Methods (Gourieroux, Monfort, Trognon 1985).

Finally, the third objective of this paper is to discuss applications of these methods to real data from Groupe Monceau.

The paper is organised as follows. In section 2 we discuss the inconsistency problem, using several examples. In section 3 we consider standard rating models and show that they are exposed to this inconsistency risk. Section 4 provides an overview of semiparametric methods. In section 5 we propose a semiparametric framework for a priori rating models. Section 6 considers the case of a posteriori rating models. A discussion of the practical implementation of a priori rating models is presented in section 7. Section 8 concludes.

2. SOURCES OF POTENTIAL INCONSISTENCY

We know that the maximum likelihood (ML) estimator, which maximises the likelihood function $\prod_{i=1}^{n}(f(y_i/x_i;\theta)$, is asymptotically efficient but may be inconsistent if the model is misspecified, that is to say if the true p.d.f $f_o(y_i/x_i)$ does not belong to the family $\{f(y_i/x_i;\theta), \theta \in \Theta\}$ on which the likelihood function is based. In a misspecified situation the ML estimator generally converges to a pseudo true value θ^* [see Gourieroux, Monfort, Trognon 1984a] and the conditional moment of a function of interest $g(Y_i)$ given $X_i = x_i$ based on the p.d.f. $f(y_i/x_i;\theta^*)$ may be very different from the true value of this conditional moment. So the price we have to pay for efficiency is the risk of inconsistency. We first consider this efficiency v.s. robustness trade-off. Then we show that, even if the parametric approach is abandoned, a misspecification of the link between endogenous variables may generate an inconsitency problem.

2.1 The efficiency v.s. robustness trade-off

In order to illustrate the efficiency-robustness trade-off, let us discuss two simple examples. The first example considers the case of IID (Independently

Identically Distributed) variables, whereas the second one deals with a conditional model (given exogenous variables)

An I.I.D. Case

We have observed n IID positive random variables $Y_i, i = 1, \ldots, n$. We assume that the common distribution of the $Y_i's$ is log-normal, i.e. that the distribution of Log Y_i belongs to the family $N(m, \sigma^2)$. It is well-known that the Maximum Likelihood (ML) estimator of m, and σ^2 are asymptotically efficient. They are given by :

$$\begin{aligned} \hat{m}_n &= \frac{1}{n}\Sigma_{i=1}^n \text{ Log } Y_i \\ \hat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^n (\text{ Log } Y_i - \hat{m})^2 \end{aligned}$$

Now suppose that we are interested in the true expectation of Y_i (denoted by μ_o. Within the model this expectation is given by $\exp(m + \frac{\sigma^2}{2})$ and we may propose the ML estimator $\exp(\hat{m}_n + \frac{\hat{\sigma}_n^2}{2})$.

Let us now assume that the model is misspecified and that the distribution of Y_i is the standard exponential probability (which does not belong to the log-normal family). The true mean of Y_i is equal to 1, and it is easily seen that $\exp\left(\hat{m}_n + \frac{\hat{\sigma}_n^2}{2}\right)$ converges to the pseudo-true value :

$$\exp\left(m^* + \frac{\sigma^{*2}}{2}\right)$$

where m^* and σ^{*2} are the mean and the variance of Log Y_i, when Y_i is exponentially distributed. The distribution of Log Y_i is a Gumbel distribution (the density of which is $\exp(x - \exp(x))$, and this implies that the limit of interest is equal to $\exp\left(-0.577 + \frac{\pi^2}{12}\right) = 1.28$. So the ML estimator which has optimal properties if the model is well specified, may be severely inconsistent if the model is misspecified ; in the example considered here the asymptotic bias is equal to 28 %. Of course, a robust estimator of the true expectaiton of Y_i is the empirical mean $\bar{Y}$, which is always consistent. If the model is well specified the asymptotic variance of $\sqrt{n}(\bar{Y}_n - \mu_o)$ is $\exp(2m_o + \sigma_o^2)[\exp(\sigma_o^2) - 1]$, which is larger than the asymtptotic variance of $\sqrt{n}[\exp(\hat{m}_n + \frac{\hat{\sigma}_n^2}{2}) - \mu_o]$, which is equal to $\exp(2m_o + \sigma_o^2)[\sigma_o^2 + \frac{\sigma_o^4}{2}]$. So the trade-off is clearly explicited.

A conditional model

We postulate that, conditionally to positive exogenous (IID) variables x_i, the endogenous positive variables y_i are independently distributed and that Log Y_i follows $N(a + bx_i, \sigma^2)$. The ML estimators $\hat{a}_n, \hat{b}_n, \hat{\sigma}_n^2$ of a b and σ^2 are given by the ordinary least squares method. Now let us assume that the true distribution is indeed lognormal but with heteroscedasticity ; more precisely let us assume that the true distribution of Log Y_i is $N(a_o + b_o x_i, c_o + d_o x_i)$ (with $c_o > 0, d_o > 0$). It is readily seen that $\hat{a}_n$ converges to a_o, $\hat{b}_n$ converges to b_o and $\hat{\sigma}_n^2$ converges to $c_o + d_o E X_1$.

If we are interested in the mean effect of X_i on Y_i, measured by the true conditional expectation

$$E_o\left(Y_i/X_i = x_i\right) = \exp\left[a_o + \frac{1}{2}c_o + (b_o + \frac{1}{2}d_o)x_i\right],$$

we will propose the ML estimator $\exp\left(\hat{a}_n + \hat{b}_n x_i + \frac{1}{2}\hat{\sigma}_n^2\right)$ which converges to $\exp[a_o + \frac{1}{2}(c_o + d_o E X_1) + b_o x_i]$, and the estimator is inconsitent. In particular the coefficient of x_i, which is crucial in a scoring problem, will be badly estimated (asymptotically underestimated here since d_o is positive); similarly the estimator of the intercept will be also inconsistent (and asymptotically overestimated).

2.2 Misspecification of the links between endogenous variables.

Let us assume that Y_i is made of N_i, the number of claim, and (S_{ik}) the size of the claims. We are interested in the expectation of the total claim amount :

$$C_i = \sum_{k=1}^{N_i} S_{ik}$$

If we assume that N_i and (S_{ik}) are independent and that the S_{ik} are IID we have :

$$\begin{aligned} m_C = EC_i &= E[E(C_i/N_i)] \\ &= m_S E(N_i) \\ &= m_S m_N \end{aligned}$$

where $m_S = ES_{ik}$ and $m_N = EN_i$.

So we may estimate m_C by $\bar{N}\bar{S}$ where $\bar{N}$ is the empirical mean of the N_i and $\bar{S}$ the empirical mean of the S_{ik}.

Now if, in fact, N_i and (S_{ik}) are independent conditionally to a positive latent variable U_i and if $E(N_i/U_i) = \lambda U_i, E(S_{ik}/U_i) = \mu U_i$, we have, using the notation m_U and σ_U^2 for the mean and the variance of U_i :

$$\begin{aligned} m_C = E(C_i) &= E\left[E(C_i/N_i, U_i)\right] \\ &= \mu E\left(N_i U_i\right) \\ &= \mu E\left[U_i E(N_i/U_i)\right] \\ &= \lambda\mu\left(m_U^2 + \sigma_U^2\right) \end{aligned}$$

The estimator $\bar{N}\bar{S}$ of m_C converges to :

$$EN_i ES_i = \lambda\mu m_U^2$$

and therefore is asymptotically biased. More precisely $\bar{N}\bar{S}$ understimate m_C, the asymptotic bias being $\lambda\mu\sigma_U^2$.

3. THE PARAMETRIC NATURE OF THE CLASSICAL RATING MODELS .

As shown in this section all the classical rating models are parametric and, therefore, exposed to the inconsistency risk.

3.1 Models based on frequency

The more popular model based on frequency is the Poisson model, in which the number of claims N_i, for which policyholder i is responsible during a given period, is assumed to follow the Poisson distribution $\mathcal{P}[\exp(x_i'\theta)]$, where x_i is a vector of exogenous variables (for sake of notational simplicity we omit the time index). It turns out that, fortunately, the Poisson model is robust to specification errors, in the sense that the ML procedure based on this model is consistent and asymptotically normal even if the true distribution is not Poisson, provided that the conditional expectation $E_\theta[N_i/x_i] = \exp(x_i'\theta)$ is well specified. However care is needed in the computation of the asymptotic variance-covariance matrix and, therefore, in the testing procedures (see section 4).

Moreover the a priori rating based on the Poisson model is sometimes completed by an a posteriori rating in which the conditional distribution of N_i given x_i and an individual unobservable (positive) effect U_i is assumed to follow the Poisson distribution $\mathcal{P}[U_i \exp(x_i'\theta)]$. Since U_i is unobservable it has to be integrated out, in order to obtain the conditional distribution of N_i given x_i only. At this stage, the distribution of U_i is usually assumed to be Gamma in order to get a tractable conditional distribution of N_i given x_i, namely a Negative Binomial distribution, and, consequently, a tractable likelihood function. It is clear that the latter assumption is rather arbitrary and that the misspecification risk is high. In the semi nonparametric a priori rating models, the a posteriori component is often assumed parametric in order to have the opportunity to use simulation based econometric methods (see Gourieroux-Monfort 1996 and 1997).

3.2 Models based on frequency and severity.

The misspecification and, therefore, inconsistency risk is higher in the models based on frequency and severity because they usually rest on a long list of untested assumptions. In the standard approach it is assumed that :

- the number of claims N_i, is distributed as $\mathcal{P}[\exp(x_i'b)]$
- the sizes of the claims S_{ik} are independently log-normally distributed
- the distribution of Log S_{ik} is $N[x_i'c, \sigma^2]$ in particular, this distribution is homoscedastic
- N_i and the (S_{ik}) are independent, conditionally to the exogenous variables.

In this model the total claim amount is :

$$C_i = \sum_{k=1}^{N_i} S_{ik}$$

and the pure premium is :

$$\begin{aligned} E(C_i/x_i) &= EN_i ES_{ik} \\ &= \exp\left[x_i'(b+c) + \frac{\sigma^2}{2}\right] \end{aligned}$$

Clearly this formula heavily depends on the assumptions made above, in particular log-normality, homoscedasticity and independence (see section 2.1. above).

The situation is even more serious in a posteriori modeling based on frequency and severity. In this kind of approach additional ad hoc assumptions are made about the distribution of heterogeneity term introduced in the modeling of S_{ik} (for instance inverse gamma in Frangos-Vrontos (2001)) or Log S_{ik} (for instance Gaussian in Pinquet (2001)) or in the modelling of the conditional behavior of C_i given N_i (Gamma in Gourieroux (1999 Chapter 8)).

4. AN OVERVIEW OF RELEVANT SEMIPARAMETRIC INFERENCE METHODS

4.1 The basic problem

The econometric literature has proposed a large set of semiparametric methods which do not necessitate assumptions about the probability distributions in order to get a CAN (Consistent Asymptotically Normal) estimator of the parameter θ characterising conditional moments. In this section we propose a brief summary of the methods which are relevant for solving the basic problem of rating models. This basic problem is to estimate pure premia and therefore to estimate the true conditional expectations $E_o(C_i/x_i)$. So, in the a priori modelling, the only assumption needed is to postulate that this unknown functin of x_i belongs to some family $m(x_i, \theta)$, where $m(.)$ is known and θ is an unknown p-dimensional parameter. The case of a posteriori modelling will be discussed in section 6.

4.2. Extremal estimators

All the relevant estimators will be extremal estimators. An extremal estimator $\hat{\theta}_n$ is, by definition, obtained by maximizing and an objective function $Q_n(Y, X, \theta)$ where $Y = (Y_1, \ldots, Y_n), X = (X_1, \ldots, X_n)$. Under technical conditions this estimator is consistent if the limit function $Q_\infty(\theta) = \lim_{n\to\infty} Q_n(Y, X, \theta)$ exists and has a unique maximum at the true value θ_o (see Gourieroux-Monfort chapter 24). Moreover $\sqrt{n}(\hat{\theta} - \theta_o)$ is CAN and the asymptotic variance-covariance matrix is $\Sigma(\theta_o) = J^{-1}(\theta_o)I(\theta_o)J^{-1}(\theta_o)$ where $J(\theta_o) = -\dfrac{\partial^2 Q_\infty(\theta_o)}{\partial\theta\partial\theta'}$ and $I(\theta_o)$ is the asymptotic variance-covariance matrix of $\dfrac{1}{\sqrt{n}} Q_n$.

4.3. Pseudo Maximum Likelihood (DML) Methods

These methods have been developed by Gourieroux, Monfort and Trognon (1984 a and b). We focus here on one of these methods : the pseudo maximum likelihood method of order 1 (PML1). In this approach the only assumption made is precisely the one we are interested in, i.e. that $E_o(y_i/x_i)$ belongs to a given family $m(x_i, \theta)$, that is to say that $E_o(y_i/x_i) = m(x_i, \theta_o)$ for a unique θ_o. The main idea is to find a family $f(y, m)$ of probability density functions (p.d.f.) indexed by their mean m, such that the pseudo likelihood function $\prod_{i=1}^n (y_i, m(x_i, \theta))$ based on this family provides, when maximized, a CAN estimator $\hat{\theta}_n$ of θ_o.

The basic result is that this result holds if, and only if, $f(y, m)$ is a linear exponential family, that is if it can be written :

$$f(y, m) = \exp[A(m) + B(y) + C(y)m]$$

for some functions A,B,C.

This condition is true for some standard families, in particular the Gaussian family $N(m, \sigma_o^2)(\sigma_o^2$ fixed at any positive value), the Poisson family $\mathcal{P}(m)$ and the Gamma family $\gamma(\alpha_o, \dfrac{m}{\alpha_o})$ (where α_o is fixed at any positive value). The objective functions in these three cases are respectively :

$$-\sum_{i=1}^{n}\left[y_i - m(x_i,\theta)\right]^2/\sigma_o^2$$

$$\sum_{i=1}^{n}\left[y_i \text{ Log } m(x_i,\theta) - m(x_i,\theta)\right]$$

$$-\sum_{i=1}^{n}\left[\frac{y_i}{m(x_i,\theta)} + \text{ Log } m(x_i,\theta)\right]\alpha_o$$

Note that the values σ_0^2 and α_0 are irrelevant in the maximisation and can therefore be dropped; also note that in the PML method associated with the Poisson and the Gamma familier $m(x_i,\theta)$ must be positive.

All these methods provided CAN estimators, whatever the true distribution is. The asymptotic variance covariance matrices of the estimators must, however, be computed with the formula of section 4.2 (sometimes called the robust formula) and not with the inverse of the information matrix.

It is seen, in particular, that, as announced previously, the Poisson family is robust with respect to misspecifications on the distribution.

4.4 Generalized Method of Moments (GMM)

This method has been proposed by Hansen (1982) and we specify it in the context of interest here.

The basic assumption of existence and uniqueness of a θ (denoted θ_o) satisfying $E_o(y_i/x_i) = m(x_i,\theta)$ can be reformulated as existence and uniqueness of a θ satisfying :

$$E_o\{A(x_i)[y_i - m(x_i,\theta)]\} = 0$$

for a K-vector $A(x_i)$ of "instruments" (with K $\geq$ p). When the number K of instruments is equal to the number of parameters (a condition which is not restrictive, see 4.5) the GMM reduces to minimizing :

$$||\sum_{i=1}^{n} A(x_i)[y_i - m(x_i,\theta)]||^2$$

where $||.||$ is the usual norm in $I\!R^p$.

or to solve :

$$\sum_{i=1}^{n} A(x_i)\left[y_i - m(x_i,\theta)\right] = 0$$

The estimator thus obtained is CAN and its asymptotic variance-covariance matrix is given by the formula of section 4.1 (see Gourieroux-Monfort, chapter 9, for more details).

It turns out that, in this case, the asymptotic variance covariance matrix is :

$$\left\{E_{X_1}\left[\frac{\partial m(X_1,\theta_o)}{\partial\theta}A'(X_1)\right]\left\{E_{X_1}\left[A(X_1)V_o(Y_1/X_1)A'(X_1)\right]\right\}^{-1}E_{X_1}\left[A\frac{\partial m}{\partial\theta'}(X_1,\theta_o)\right]\right\}^{-1}$$

The semiparametric efficiency bound

It can be shown (see Gourieroux-Monfort (1996 chapter 23)) that all the semiparametric estimators of θ_o, and in particular the PML1 and GMM estimators introduced above, have an asymptotic variance-covariance matrix which is larger than a semiparametric efficiency bound given here by :

$$\left\{E_{X_1}\left[\frac{\partial m(X_1,\theta_o)}{\partial\theta}V_0^{-1}(Y_1/X_1)\frac{\partial m(X_1,\theta_o)}{\partial\theta'}\right]\right\}^{-1}$$

An obvious question is : how to reach this bound ? As far as the GMM method is concerned, comparing this bound with the formula of the asymptotic variance-covariance matrix given in section 4.4, we see that the optimal instruments are given by :

$$A^*(X_i) = \frac{\partial m(X_i,\theta_o)}{\partial\theta}V_o^{-1}(Y_i/X_i)$$

However these instruments are not feasible since θ_o is unknown (but we could replace it by a consistent estimator) and, more importantly, $V_o(Y_i/X_i)$ is unknown and not specified. If we want to go further we could assume that $V_o(Y_i/X_i)$ belong to some parametric family, but this would contradict our robust approach, or we could estimate $V_o(Y_i/X_i)$ by nonparametric kernel approaches but, in practice, we would face the "curse of dimensionality" problem.

A pragmatic approach consists in choosing a set of instruments which is likely not too far from the optimal one and which leads to easily computable and easily interpretable estimators. It is important to stress that, in any case, the estimators are CAN.

The attainability of the bound in the PML framework can also be discussed by using the Quasi Generalized PML approach [see Gourieroux-Monfort-Trognon 1984 a]. In particular it is easily that the three PML1 methods considered above are optimal in the following cases :

$V_o(Y_1/X_1)$ constant for the Gaussian case.
$V_o(Y_1/X_1)$ proportional to $m^2(X_1,\theta_o)$ in the Gamma case

Moreover it is easily seen that in the special case where $m(X_1,\theta) = \exp(X_1'\theta)$, the PML1 method based on the Poisson family is optimal if $V_o(Y_1/X_1)$ is proportional to $m(X_1,\theta_0)$.

4.6 Asymptotic Least Squares (ALS)

This method has been proposed by Gourieroux-Monfort-Trognon (1985). It allows for estimating and testing an auxiliary parameter λ defined from θ by nonlinear restriction $b(\theta,\lambda)=0$. Here we restrict ourselves to the case where the previous constraints are :

$$\theta = S\lambda$$

where S is a matrix (made of 0 and 1) such that each $\theta_j = S_j\lambda$ is either 0 or a component of λ. In other words λ is obtained from θ by putting some components to zero.

The ALS method provides, in this case, a very convenient method for testing hypotheses of the form $\theta = S\lambda$ and estimating λ, in any of the semiparametric methods proposed above. Denoting by $\hat\theta_n$ the unconstrained estimator and $\hat\Sigma$ a consistent estimator of the asymptotic variance-covariance matrix $J^{-1}(\theta_o)I(\theta_o)J^{-1}(\theta_o)$ (see section 4.2), an equivalent model is the linear model $\hat\theta_n = \theta + u$ where u is zero-mean and has a covariance matrix equal to $\frac{\hat\Sigma}{n}$, or equivalently :

$$
\begin{aligned}
M\hat{\theta}_n &= M\theta + v \\
\text{with} \quad Ev &= 0, V(v) = I, MM' = \frac{\hat{\Sigma}}{n}
\end{aligned}
$$

Constrained estimation of θ and tests on θ can be performed in this "asymptotic" model, in the usual way (t tests, or Wald tests) and are asymptotically equivalent to the corresponding constrained estimation and tests within the semiparametric procedure used for obtaining $\hat{\theta}_n$.

The huge advantage of this method is that the estimation or testing procedures in the linear model are considerably faster than their equivalent in the initial semiparametric framework (PML or GMM here).

5. ROBUST INFERENCE FOR A PRIORI RATING MODEL

5.1 A multiplicative specification

In order to try to avoid the misspecification problems discribed in section 2, it is natural to specify only the function of interest, namely the conditional expectation $E_o(C_i/x_i), C_i$ being the total claim amount and x_i a vector of exogenous variables. We assume here that $E_o(C_i/x_i)$ belongs to the family $\{\exp(x_i'\theta), \theta \in I\!R^p\}$ in order to have a usual multiplicative formula. In other words, the function $m(x_i, \theta)$ introduced in section 4 is $\exp(x_i'\theta)$

5.2 Semiparametric inference

We know that all the methods proposed in section 4 provide CAN estimators of θ_o, the true value of θ. However we might want to go further in the direction of optimality.

From the results of section 4, we know that the unfeasible optimal instruments in the GMM approach are :

$$x_i \exp(x_i'\theta_o) V_o^{-1}(Y_i/x_i)$$

where $V_o(Y_i/X_i)$ is the true conditional variance of Y_i given X_i. This optimal set of instruments is unknown because θ_o and $V_o(Y_i/x_i)$ are unknown. The true value θ_o can be replaced by any first stage consistent estimator $\tilde{\theta}_n$. As far as $V_o(Y_i/X_i)$ is concerned, we might impose a parametric specification, for instance $\exp(x_i'\varphi)$, and estimate φ by a semiparametric method applied to $(C_i - \exp(x_i'\tilde{\theta}_n))^2$. However this approach increases the set of assumptions, and moreover, it is known that two stage procedures may have bad finite sample properties. A more pragmatic approach consists in considering important particular specifications leading to one stage procedures. Three particular cases are natural :

$A : V_o(Y_i/x_i)$ constant

$B : V_o(Y_i/X_i$ proportional to $m(x_i,\theta_o)\exp(x_i'\theta_o)$

$C : V_o(Y_i/X_i)$ proportional to $m^2(x_i,\theta_o) = \exp(2x_i'\theta_o)$

In case A the gaussian PML1 method i.e. the nonlinear least squares (NLS) method, reaches the semiparametric efficiency bound ; this method consist in minimizing $\sum_{i=1}^{n}(Y_i - \exp(x_i'\theta))^2$. The GMM method would lead to a two stage method with the instruments $x_i\exp(x_i'\tilde{\theta}_n)$ and is not recommenced.

In case B, the optimal instruments are simply x_i and therefore a one step method in avalaible. Moreover this method is equivalent to the Poisson PML1 method. It is also interesting to note that these methods are asymptotically equivalent to the weighted nonlinear least square method (WNLS) in which we minimise :

$$\sum_{i=1}^{n}\frac{(Y_i - \exp(x_i'\theta))^2}{\exp(x_i'\tilde{\theta}_n)}$$

In case C the optimal instruments are $x_i\exp(-x_i'\tilde{\theta}_n$ and necessitate a two stage procedure. Fortunately this method is equivalent to the one stage Gamma PML1 method. Note that it is also equivalent to the WNLS method in which we minimise :

$$\sum_{i=1}^{n}\frac{(Y_i - \exp(x_i'\theta))^2}{\exp(2x_i'\tilde{\theta}_n)}$$

The range of the relative weights is larger than in the previous case.

So in all the cases, a one stage method is available.

A natural benchmark method is method B, which is based on intermediate assumptions on the conditional variance and which is implicitely assuming an intermediate range of relative weights in the WNLS method.

Moreover the GMM method applied in case B) implies

$$\sum_{i=1}^{n} x_i(Y_i - \exp(x_i'\hat{\theta}_n)) = 0$$

So if x_i is made of dummy variables corresponding to various classes of a qualitative variable, we have for any such class

$$\sum_{i \in I} y_i = \sum_{i \in I} \exp(x_i'\hat{\theta}_n)$$

where I is the set of policyholders belonging to this class. In other words the rating is fair for all these classes.

These properties are summarized in the following array :

Semiparametric method	Optimality condition on $V_o(Y_i/x_i)$	Optimal GMM instruments	Optimal PML1 method	Optimal weights in WNLS
A	Constant	$x_i \exp(x_i'\tilde{\theta}_n)$	Gaussian	1
B	Proportional to $\exp(x_i'\theta_o)$	x_i	Poisson	$\exp(-x_i'\tilde{\theta}_n)$
C	Proportional to $\exp(2x_i'\theta_o)$	$x_i \exp(-x_i'\tilde{\theta}_n)$	Gamma	$\exp(-2x_i'\tilde{\theta}_n)$

5.3 Wald test based on Asymptotic Least Squares

Once parameter θ appearing in $\exp(x_i'\theta)$ has been estimated by one of the semiparametric methods proposed above, we have a consistent estimator $\hat{\theta}_n$ such $\sqrt{n}(\hat{\theta}_n - \theta_o)$ is asymptotically distributed as $N[0, \Sigma(\theta_o)]$ where $\Sigma(\theta_o) = J^{-1}(\theta_o)I(\theta_o)J^{-1}(\theta.)$ If $\hat{\Sigma}_n$ is a consistent estimator of $\Sigma(\theta_o)$ we can easily test any null hypothesis on θ_o, using the asymptotic linear model described in 4.6. So we can propose a fast downward variable selection procedure. Starting from a maximal set of possible variables, we reject at each the one which is the less significant. Note that each variable may appear through several elementary variables (dummies in the qualitative case, powers of the variable in the quantitative case) and that, in this case, we test the simultaneous irrelevance of all these variables. A natural rejection criterion at each step is the highest p-value, the p-value being the minimal probability of wrong acceptation we have to impose in order to accept the variable (the acceptation of variable meaning the rejection of the hypothesis of nullity of the corresponding components of θ). We stop the procedure when all the remaining p-values are smaller than a given threshold, for instance 5%.

Even if the number of tests is high this procedure is fast because we can use standard tests of the linear model. This would not be true in an upward procedure.

6. SEMIPARAMETRIC A POSTERIORI RATING

6.1. The setup

The endogenous variable of interest for policyholder i and period t is denoted by Y_{it}; it might be a number of claims or a total claim amount. For each policyholder i we introduce a positive latent variable U_i measuring an individual unobservable effect and exogenous variables x_{it}. We denote by n the number of policyholders, by T the number of periods (assumed independent of i for sake of notational simplicity) and by Y_i the vector $(Y_{i1}, \ldots, Y_{iT})$.

We assume that the $U_i's$ are IID and independent of the exogenous variables. We also assume that, conditionally to the $U_i's$ and the exogenous variables, the Y_{it} are independent and we specify the conditional means and variances as follows :

$$\begin{aligned} E(Y_{it}/x_{it}, u_i) &= u_i \lambda_{it} \\ V(Y_{it}/x_{it}, u_i) &= a(u_i) \tilde{\mu}_{it} \end{aligned}$$

where λ_{it} and $\tilde{\mu}_{it}$ are parametric functions of x_{it}, for instance $\exp(x'_{it}b)$ and $\exp(x'_{it}\tilde{c})$ respectively and $a(u_i)$ is some positive function. For obvious identification reasons we can always assume that $E(U_i) = 1$ and we denote $V(U_i)$ by σ^2.

Under these assumptions we have :

$$
\begin{aligned}
E(Y_{it}/x_{it}) &= \lambda_{it} \\
V(Y_{it}/x_{it}) &= V(U_i\lambda_{it}) + E[a(U_i)\tilde{\mu}_{it}] \\
&= \sigma^2\lambda_{it}^2 + \mu_{it}
\end{aligned}
$$

where $\mu_{it} = \tilde{\mu}_{it}E[a(U_i)]$ (the constant $E[a(U_i)]$, which may be non identifiable, is incorporated in the specification of $\tilde{\mu}_{it}$; if $\tilde{\mu}_{it} = \exp(x'_{it}\tilde{c})$ this reduces to a change of intercept and we note $\mu_{it} = \exp(x'_{it}c)$) and :

$$
\begin{aligned}
cov(Y_{ir}, Y_{is}/x_{ir}, x_{is}) &= E\{cov[(Y_{ir}, Y_{is})/x_{ir}, x_{is}, u_i]/x_{ir}, x_{is}\} \\
&+ cov\{[E(Y_{it}/x_{ir}, u_i), E(Y_{is}/x_{is}, u_i]/x_{ir}, x_{is}\} \\
&= cov\ [(U_i\lambda_{ir}, U_i\lambda_{is})/x_{ir}, x_{is}] \\
&= \sigma^2\lambda_{ir}\lambda_{is}
\end{aligned}
$$

Applying the particular cases A,B,C considered in 5.2 to the conditional variance $V(Y_{it}/x_{it}, u_i)$, we obtain three particular specifications for $\mu_{it} = E[V(Y_{it}/x_{it}, u_i)]$:

$$
\begin{aligned}
A\ : \mu_{it} &= c \\
B\ : \mu_{it} &= cE(Y_{it}/x_{it}, u_i) = c\lambda_{it} \\
C\ : \mu_{it} &= \tilde{c}[E(Y_{it}/x_{it}, u_i)]^2 = \tilde{c}\lambda_{it}^2 EU_i^2 \\
&= c\lambda_{it}^2(\text{say})
\end{aligned}
$$

6.2. Estimation

Since $E(Y_{it}/x_{it}) = \lambda_{it}$, the parameters appearing in λ_{it} (for instance b in $\exp(x'_i b)$) can be consistently estimated by any semiparametric method proposed in section 4, applied to Y_{it}. We get estimators $\hat{\lambda}_{it}$ of λ_{it}.

σ^2 can be consistently estimated, for instance, by the mean $\hat{\sigma}^2$ of empirical covariances between $\dfrac{Y_{ir}}{\hat{\lambda}_{ir}}$ and $\dfrac{Y_{is}}{\hat{\lambda}_{is}}$, $s \neq r$.

Finally the parameters appearing in μ_{it} can be consistently estimated by any semiparametric method of section 4 applied to $Y_{it}^2 - \hat{\lambda}_{it}^2(1+\hat{\sigma}^2)$

6.3. Optimal Bonus-Malus coefficients

Since we do not have made assumptions about the probability distributions we cannot compute the conditional expectations $E(Y_{i,T+1}/Y_i, x_{i1}, \ldots, x_{i,T+1})$ but we can compute the best prediction linear in Y_i. For sake of notational simplicity we now omit subscript i and the conditioning with respect to the exogenous variables. The best linear prediction of Y_{T+1} given Y will be denoted by $E^L(Y_{T+1}/Y)$.

We have :

$$E^L(Y_{T+1}/Y) = EY_{T+1} + cov(Y_{T+1}, Y)[V(Y)]^{-1}(Y - EY)$$

with :

$$\begin{aligned} EY_{T+1} &= \lambda_{T+1} \\ EY &= (\lambda_1, \ldots, \lambda_T)' = \lambda \\ cov(Y_{T+1}, Y) &= \sigma^2 \lambda_T \lambda' \\ V(Y) &= \sigma^2 \lambda \lambda' + diag(\mu) \end{aligned}$$

where $\mu = (\mu_1, \ldots, \mu_T)'$.

It is easily seen that :

$$[V(Y)]^{-1} = diag(\mu^{-1}) - \frac{\sigma^2 (\lambda\mu^{-1})(\lambda\mu^{-1})'}{1 + \sigma^2 \parallel \lambda\mu^{-1/2} \parallel^2}$$

where $\mu^{-1}, \lambda\mu^{-1}, \lambda\mu^{-1/2}$ are the vectors whose components are, respectively, $\mu_t^{-1}, \lambda_t\mu_t^{-1}, \lambda_t\mu_t^{-1/2}$.

This implies :

$$E^L(Y_{T+1}/Y) = \lambda_{T+1} \frac{1 + \sigma^2 (\lambda\mu^{-1})' Y}{1 + \sigma^2 \parallel \lambda\mu^{-1/2} \parallel^2}$$

or more, explicitely :

$$E^L(Y_{T+1}/Y) = \lambda_{T+1} \frac{1 + \sigma^2 \sum_{t=1}^{T} \lambda_t \mu_t^{-1} Y_t}{1 + \sigma^2 \sum_{t=1}^{T} \lambda_t^2 \mu_t^{-1}}$$

Since

$$\begin{aligned}
E^L(Y_{T+1}/Y) &= E^L[E(Y_{T+1}/Y)] \\
&= E^L\{E[E(Y_{T+1}/Y, U)/Y]\} \\
&= \lambda_{T+1} E^L[E(U/Y)] \\
&= \lambda_{T+1} E^L(U/Y)
\end{aligned}$$

the best linear prediction of U, is :

$$E^L(U/Y) = \frac{1 + \sigma^2 \sum_{t-1}^{T} \lambda_t \mu_t^{-1} Y_t}{1 + \sigma^2 \sum_{t=1}^{T} \lambda_t^2 \mu_t^{-1}} = BM$$

This quantity denoted by BM is the semiparametric optimal Bonus Malus coefficient. In the case A,B,C considered above this coefficient becomes :

$$A \ : BM \ = \ \frac{1 + \sigma^2 c^{-1} \sum_{t=1}^{T} \lambda_t Y_t}{1 + \sigma^2 c^{-1} \sum_{t=1}^{T} \lambda_t^2}$$

$$B \ : BM \ = \ \frac{1 + \sigma^2 c^{-1} \sum_{t=1}^{T} Y_t}{1 + \sigma^2 c^{-1} \sum_{t=1}^{T} \lambda_t}$$

$$C \ : BM \ = \ \frac{1 + \sigma^2 c^{-1} \sum_{t=1}^{T} \lambda_t^{-1} Y_t}{1 + T\sigma^2 c^{-1}}$$

Note that if Y_t is the number of claims and if the conditional distribution of Y_t given (x_t, u) is assumed to be Poisson we have :

$$V(Y_t/x_t, u) = E(Y_t/x_t, u) = u\lambda_t \text{ and } \mu_t = \lambda_t$$

so, we are in case B with $c = 1$ and we get

$$BM = \frac{1 + \sigma^2 \sum_{t=1}^{T} Y_t}{1 + \sigma^2 \sum_{t=1}^{T} \lambda_t}$$

This BM coefficient is also the one obtained in the parametric case where, moreover, U is assumed to follow the distribution $\gamma(\frac{1}{\sigma^2}, \frac{1}{\sigma^2})$ [see Dionne-Vanasse (1992) and Gourieroux (1999)]. So, in some sense, this BM coefficient is robust with respect to the Gamma assumption about the distribution of U but it is not robust with respect to the Poisson assumption about the conditional distribution of Y_t given (x_t, u). The robustness of the Poisson assumption in a priori rating is lost in a posteriori rating.

Going back to the general setup, it is worth noting that, for the true value u of U and applying the strong law of large numbers to the variables $\lambda_t \mu_t^{-1} Y_t$ and $\lambda_t^2 \mu_t^{-1}$ we get :

$$\frac{1}{T}\sum_{t=1}^{T} \lambda_t \mu_t^{-1} Y_t \underset{T\to\infty}{\longrightarrow} uE\lambda_1^2 \mu_1^{-1}$$

$$\frac{1}{T}\sum_{t=1}^{T} \lambda_t^2 \mu_t^{-1} \underset{T\to\infty}{\longrightarrow} E\lambda_1^2 \mu_1^{-1}$$

and

$$BM = \frac{\frac{1}{T} + \sigma^2 \frac{1}{T}\sum_{t=1}^{T} \lambda_t \mu_t^{-1} Y_t}{\frac{1}{T} + \sigma^2 \frac{1}{T}\sum_{t=1}^{T} \lambda_t^2 \mu_t^{-1}} \underset{T\to\infty}{\longrightarrow} u$$

Therefore, the BM coefficient converges to the value taken by the latent variable U.

7.PRACTICAL IMPLEMENTATION

The a priori rating methods presented above has been applied to car insurance data sets from Groupe Monceau. This application has shown that, in addition to their theoretical robustness proved in the previous sections, these methods are able to tackle the practical problems usually encountered. Let us discuss some of these problems.

Duration

It is important to define, for each contract, a set of spells within which the characteristics of this contact (the car characteristics in particular) do not change. So we have to work with spells of different lengths and the basic observation i becomes a spell of a given contract. If d_i is the duration of this spell (using for instance the year as time unit) the conditional expectation $E_o(C_i/x_i)$ is specified as $d_i \exp(x_i'\theta)$ or $\exp(\text{Log } d_i + x_i'\theta)$, and the semiparametric methods are easily generalised.

Selection

The selection of exogenous variables presented in 5.3, may be based on a partition of elementary variables, each set of this partition containing either dummies associated with a qualitative variable or powers of a quantitative variable. The ALS method is able to treat this case.

Discretisation

The discretisation of a quantitative variable may be based, as usual, on its empirical distribution but, also, on the profile of its risk factor defined as the exponential of the polynomial used in the selection procedure.

Agregation

For qualitative variables, or discretised quantitative variables, it may be useful to agregate the classes of the partition. For this purpose, the selection procedure based on ASL is easily adapted to the case of testing equality of coefficients.

Interaction

Lack of interaction is characterised by the nullity of some coefficients ; for instance in the case of two quantitative variables the lack of interaction may be characterised by the nullity of the coefficients of the cross terms in the polynomial. Therefore this type of test may also be incorporated in the ALS based selection procedure.

Outliers

Like all statistical methods, the GMM may be sensitive to outliers and we may have to trim large claims, that are, in motor insurance, liability claims with serious body injuries or deaths. Trimming these claims (at the level of 100. 000 € for example) may hide the actual risk exposition of some categories of risk (like young drivers). There is no other solution than to do a "manual" analysis and to add a "manual" correction to the results of the method.

Data extraction and evaluation

Obtaining from the IT Department of the insurance compagnies the right data you need for your computations is not always easy : you need detailed data with one record for each contract and one record for each claim during the observation periods. That means some new work for the IT departments, which have always higher priorities... Afterwards, when you receive the data, you have to check them, to verity the exactitude and the completion of the extractions. There is often some work to do there.

Marketing aspects linked with an innovative tariff

Implementing such a new tariff has marketing consequences : the fact you make tariff innovations and go away from the generally previously practiced tariffs, makes you competitive, sometimes very competitive, in new layers of customers, you didn't know at all before. A usual problem is the common and unavoidable assumption that the new risks you underwrite have the same behaviour than the risks you have in your portfolio. This is not necessary true. When your development is not too fast, you can correct and adjust the tariff with the reality of the new risks. Otherwise you can be in a dangerous situation. Another problem is the use you make in the tariff of items that had no tariff impact before : for example the number of authorised drivers. As long as information had no consequences on the premium, the information you obtained was exact. But, your sales force understands quickly that, for example, with two authorised drivers, the premium is less high than with just one driver (because you observed on your previous statistics that the risk with two drivers, married people, was less high than this with just one driver) and you will have all new businesses with two authorized drivers...

Subsequent tariff updating

Another problem lies in the updating you have to do for this tariff, in order to take the last risk statistics into account. The structure of such a tariff, based on the multiplication of coefficients, with a relatively large number of variables (commonly 10 to 20) has a consequence, when you update the coefficients : some risks can see their premium becoming significantly higher or smaller. Even if you can calibrate the premium you ask in a renewal of contract for clients already in portfolio, for new businesses, when you have sales forces that position their sales efforts in particular directions, they can be disoriented when your updated tariff makes suddenly specific categories of insured unattainable, and other categories, ignored until this moment, accessible. Sales forces do not like changes.

These disadvantages of such a "mathematical tariff", you update continuously with the reality of the risk, do not exist with direct selling.

8. CONCLUSION

The methods proposed in this paper are promising since they feature both theoretical robustness properties and practical flexibility. They can be used for a priori, a posteriori, or mixed rating models. The a posteriori aspect has not yet be experimented on real data and could been an interesting topic for further research.

REFERENCES

Chamberlain G. (1987) :" Asymptotic Efficiency in Estimation with Conditional Moment Restrictions" , Journal of Econometrics, 34, 305-334.

Dionne G. and Vanasse C. (1992) : "Automobile Insurance Ratemaking in the Presence of Asymmetrical Information", Journal of Applied Econometrics, 7, 149-165.

Frangos N.E. and Vrontos S.D. (2001) : "Design of Optimal Bonus-Malus Systems with a Frequency and Severity Component on an Individual Basis in Automobile Insurance", ASTIN Bulletin, 31, n°1, 5-26.

Gourieroux C. (1999) : "Statistique de l'assurance", Economica, Paris

Gourieroux C. Monfort A. and Trognon A. (1984 a) : "Pseudo Maximum Likelihood Methods Theory", Econometrica, 52, 681-700.

Gourieroux C. Monfort A. and Trognon A. (1984b) : "Pseudo Maximum Likelihood Methods : Applications to Poisson Models", Econometrica, 52, 700-720.

Gourieroux C. and Monfort A. (1985) : "Moindres ccarrés asymptotiques", Annales de l'INSEE, 58, 91-122.

Gourieroux C. and Monfort A. (1995) : "Statistics and Econometric Models", 2 volumes, Cambridge University Press.

Gourieroux C. and Monfort A. (1996) : "Simulation Based Econometric Methods", Oxford University Press.

Gourieroux C. and Monfort A. (1997) : "Modèles de comptage semi-paramétriques", in Econométrie Appliquée, Economica - Presses des HEC de Montréal.

Hansen L.P. (1982) : "Large Sample Theory of Generalized Method of Moments Estimators", Ecometrica, 50, 1029-10254.

Monfort A. (2002) : "Variables latentes et modélisation statistique en assurance", Document CREST.

Newey W.K. (1988) : "Adaptative Estimation of Regression Models via Moment Restrictions", Journal of Econometrics, 38, 301-309.

Newey W.K. (1989) : "Introduction à la Théorie des Bornes d'Efficacité Semiparamétrique", Annales d'Economie et de Statistique, 13, 1-48.

Pinquet J. (2001) : "Experience Rating Through Heterogenous Models", in Handbook of Insurance, Chapter 15, Kluwer.

White H. (1981) : "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-25.