HAL open science

Experience rating in non-life insurance

Jean Pinquet

► To cite this version:

Jean Pinquet. Experience rating in non-life insurance. 2012. hal-00677100

HAL Id: hal-00677100

https://hal.science/hal-00677100

Preprint submitted on 7 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE POLYTECHNIQUE
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

EXPERIENCE RATING IN NON-LIFE INSURANCE

Jean PINQUET

March 2012

Cahier n° 2012-10

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.enseignement.polytechnique.fr/economie/
mailto:chantal.poujouly@polytechnique.edu

EXPERIENCE RATING IN NON-LIFE INSURANCE

Jean PINQUET[1]

March 2012

Cahier n° 2012-10

Résumé: Cet article présente les modèles statistiques conduisant à la tarification a posteriori en assurance. Les corrélations entre variables de risque peuvent s'expliquer de manière endogène ou exogène. L'interprétation retenue par les modèles actuariels est exogène et reflète la contagion positive habituellement observée pour les nombres de sinistres. Cette contagion positive peut être expliquée par la révélation dans le temps de caractéristiques cachées des lois de risque. Ces caractéristiques sont représentées par des effets fixes qui sont prédits avec un modèle à effets aléatoires. Cet article aborde les problèmes d'identification de la nature de la dynamique des données d'assurance non-vie. Des exemples de prédiction sont donnés pour des modèles de comptage avec des effets aléatoires constants ou dynamiques, une ou plusieurs équations, et pour des modèles sur les nombres et les coûts d'évènements.

Abstract: This paper presents statistical models which lead to experience rating in insurance. Serial correlation for risk variables can receive endogeneous or exogeneous explanations. The interpretation retained by actuarial models is exogeneous and reflects the positive contagion usually observed for the number of claims. This positive contagion can be explained by the revelation throughout time of a hidden features in the risk distributions. These features are represented by fixed effects which are predicted with a random effects model. This article discusses identification issues on the nature of the dynamics of non-life insurance data. Example of predictions are given for count data models with a constant or time-varying random effects, one or several equations, and for cost-number models on events.

Classification :. C13, C14, C23, C25, C30.

Mots clés : Effets fixes et aléatoires, surdispersion, principe de valeur espérée, crédibilité linéaire.

Key Words : Fixed and random effects, overdispersion, expected value principle, linear credibility approach.

[1] Université Paris 10 et Département d'Economie, Ecole Polytechnique (jean.pinquet@polytechnique.edu). Jean Pinquet acknowledges financial support from the AXA "Large Risks in Insurance" Chair (France), part of the AXA Research Fund.

1 Introduction

The assessment of individual risks in non life insurance raises problems which occur in any statistical analysis of longitudinal data. An insurance rating model computes risk premiums, which are estimations of risk levels, themselves expectations of risk variables. These variables are either numbers of claims or are related to their severity (the cost of the claim, or the duration of a compensation). The risk levels assessed in this paper are the frequency of claims and the pure premium, which refers to the expected loss or to its estimation.

Experience rating in non-life insurance is almost systematic, and can be justified with two arguments.

- The first argument is actuarial neutrality. For non-life insurance data, a claimless period usually implies a reduction in frequency premium for the next periods, whereas an accident triggers an increase in the premium. Hence bonus-malus systems (i.e. no-claim discounts and increases in premium after a claim) can be justified with an actuarial neutrality argument.

- The second argument are the incentives to risk prevention created by experience rating. There is a short-term efficiency of effort in reducing non-life insurance risks,[1] and experience rating may create these incentives under conditions which are recalled later in this article. Things are different for health and life risks. These risks are related to a capital, the depletion of which is partly irreversible. Prevention efforts are inefficient in the short run, and there is a reclassification risk which makes experience rating very uncommon.[2]

The predictive ability on risks of individual histories reflects two possible interpretations. On the one hand, histories reveal an unobserved heterogeneity, which has a residual status with respect to observable information on the risk units. On the other hand, histories modify risk levels, either through incentives or through psychological effects. Tversky and Kahneman (1973) proposed an “availability bias” theory, where the subjective estimation of the frequency of an event is based on how easily a related outcome can be brought to mind. An accident may then increase the perceived risk level and consequently prevention activities. At the opposite, the “gambler’s fallacy” argument (Tversky and Kahneman, 1974) suggests that individuals will feel protected from a risk after the occurrence of a related event. In that

[1] Risk reduction applies on frequency rather than severity in most of the economic literature. Hence prevention is of the "self-protection" rather than of the "self-insurance" type, with the Ehrlich-Becker (1972) terminology.

[2] Hendel and Lizzeri (2003) mention however term life insurance contracts in the US that offer state contingent prices, where low premiums are contingent on the insured showing he is still in good health.

case, prevention activities decrease after an accident, which entails an increase in risk as for the revelation effect of unobserved heterogeneity.

Experience rating is performed in the actuarial literature through a revelation principle. Unobserved heterogeneity on risks is taken into account with mixture models, where the mixing distribution reflects the weight of unobserved information. Individual fixed effects reflect the relative risk between an individual and his peers (i.e. with the same regression components). Experience rating is obtained through the prediction of this fixed effect, which is performed from a demixing derivation. Parametric approaches can be used (see Lemaire (1995) for a survey of frequency risk models), but semiparametric derivations pioneered by Bühlmann (1967) in the actuarial literature are also very popular. Non-life insurance is thus one of the domains that has offered to Karl Pearson a posthumous revenge on Ronald Fisher.

This article is organized as follows. Section 2 describes experience rating schemes in the non-life insurance business as well as cross-subsidies between periods. Section 3 recalls the usual representations of unobserved heterogeneity by fixed and random effects models, and the experience rating strategies in relation with the type of specification of the mixing distribution (whether parametric, semiparametric, and non parametric). Section 4 present the "generalized linear models" (Nelder and Wedderburn (1972), Zeger et al. (1986)), of current use in non-life insurance rating. Section 5 discusses the nature of the dynamics in non-life insurance, a point developed in more detail by Chiappori and Salanié (2012) in connection with economic theory. Lastly, Section 6 presents examples of frequency and pure premium risk models.

2 Experience rating schemes and cross-subsidies in the non-life insurance industry

There is a trend towards deregulation in the automobile insurance industry, but bonus-malus systems are still in force in the world (either compulsory as in France, or not but used by most of the competitors as in Belgium). A bonus-malus system summarizes an event history, where events are most often claims at fault. This coefficient is updated each year, decreases after a claimless year (no-claim discount) and increases if events are reported during the year. The insurance premium is the product of the bonus-malus coefficient and of a basic premium. A bonus-malus system enforces the experience rating policy if the basic premium does not depend on the individual history. This is not the case any more in France, but the bonus-malus system provides an information available to all the competitors in the market. Reducing information rents is now the role of bonus-malus systems, more than enforcing experience rating rules.[3]

[3]Ten years ago, the European Commission sued France, arguing that the bonus-malus system distorted competition. As an answer, French authorities argued that the bonus-malus system did

Let us consider for instance the updating rules for bonus-malus coefficients in France. A new driver begins with a bonus-malus coefficient equal to one, and this coefficient is equal to 0.95 after one year if no claim at fault is reported. The coefficient is equal to $(1.25)^n$ if n claims at fault are reported during the first year, and is bounded by 3.5. The same rules are applied later to the new coefficient. Besides, there is a lower bound of 0.5 for the coefficient. If the bonus-malus coefficient is equal to 0.95, you have a five percent bonus, whereas a claim at fault entails a twenty five percent malus. In this example, the bonus-malus coefficient is roughly an exponential function of the number of claims at fault. In other countries, the average coefficient after a given number of years is usually a convex function of the number of claims. As the bonus-malus coefficient is updated from the preceding value and from the claim history in the last year, bonus-malus systems can be expressed as Markov chains (see Lemaire (1995)).

Actual bonus-malus systems always have a "crime and punishment" flavour. The events which trigger a malus are usually claims at fault. If a no-fault system is in force as in several states of the United States and in Quebec, claims at fault are often replaced in the experience rating scheme by offences against the highway safety code. You can also think of mixing the history of claims and offences in the rating structure. In the USA, insurers have direct access to records of the Motor Vehicles Division. In states with a tort compensation system (i.e. fault is determined if the accident involves a third party), insurance companies use both types of events in their experience rating schemes. A speeding ticket related to more than fifteen m.p.h. above the speed limit entails the same penalty as an accident at fault, and so does failure to stop at a traffic light, or failure to respect a stop sign. The worst offence consists in overtaking a school bus while its red lights are blinking. It is worth nine points, instead of five for the aforementioned events.

Fairness in the rating structure is made necessary because of the difficulty to maintain cross subsidies between different risk levels in a competitive setting. Hence, risk premiums are usually seen as estimations of expectations of risk variables conditional on an information available to the insurance company. A question is raised about the private or public nature of this information. Insurance companies are not forced by competition to use private information on their policyholders in their rating structure. A compulsory bonus-malus system makes this information partly public, since it provides a summary of the policyholder's behaviour which can be shown to every competitor of the insurance company.

Cross-subsidies between the periods of a contract are termed as either "back-loading" or "front-loading", depending on whether the first periods are subsidized by the following ones, or the contrary. "Back-loading" in insurance contracts may occur when the insurer extracts a rent from the policyholder based on its use of

not enforce experience rating. They finally won the case.

private information (Kunreuther and Pauly, 1985), or from the maximization of a customer's value derived from an estimated lapse behavior (Taylor, 1986).[4] In a recent study of an Australian automobile insurance portfolio, Nini and Kofman (2011) find that average risk decreases with policyholder tenure, but that the effect is entirely due to the impact of observable information. This results contradicts the theory of informational monopoly power.[5]

3 Allowance for unobserved heterogeneity by random effects models

This section does not provide a self-content presentation of such models and of their applications to experience rating. A more detailed exposition is given in Pinquet (2000). Classic references are Lemaire (1995) for parametric models, and Bühlmann, Gisler (2005) for semiparametric approaches.[6] Denuit, Maréchal et al. (2007) provide a comprehensive presentation on count data models applied to non-life insurance. We recall later the main features of fixed and random effects models applied to experience rating, and we illustrate with a basic example in non-life insurance (i.e. a frequency risk model on a single type of event). We consider a sample of risk units, and we interpret data dynamics within these units (e.g., between different periods of time series) with a revelation principle. Three levels are used in the rating model.

- The first level is an *a priori* rating model which does not allow for unobserved heterogeneity. An important assumption is that the risk variables defined within a statistical unit are independent. Hence data dynamics are only explained by the revelation of unobserved heterogeneity.
- A second level includes individual fixed effects in the *a priori* rating model. These fixed effects reflect idiosyncratic features of risk distributions that are not represented by the regression components. The independence assumption is not challenged at this level.

[4] Kunreuther and Pauly's model is derived in a no-commitment setting, with myopic consumers (i.e. those who take decisions based on the current contract). Taylor uses a multiperiod approach where the premium is the control variable in the maximization of the customer value. The model also includes an elasticity between the lapse rate and relative prices between the incumbent insurer and its competitors.

[5] At the opposite, life and health insurance products are often front-loaded, and sometimes heavily without any surrender value as is the case for long-term care insurance. Hendel and Lizzeri (2003) provide an economic analysis of front-loading in term-life insurance in the US.

[6] In most statistical problems, a parameter set has a much smaller dimension than that of the probability set it aims at describing. A parametric approach is a one-to-one map from the parameter set to the probability set. In a semiparametric setting, the parameters are related to constraints on the probabilities.

- The third level is the random effects model. The fixed effects are assumed to be outcomes of random effects. The distributions of the random effects model are mixtures of those of the *a priori* rating model. These distributions are those of a class of real individuals with the same observable information, represented by the regression components.

Experience rating is obtained from a prediction of the fixed effect (plugged multiplicatively into the expectation of the risk variable) for the next period. This prediction can either be obtained from a posterior likelihood in a parametric setting, or from constraining the shape of the predictor in a semiparametric setting. In the latter case, this shape must be affine in order to make derivations tractable, and this type of risk prediction is usually termed as the linear credibility approach. Risk prediction with the random effects model implicitly supposes that the dynamics observed on the data are only due to a revelation mechanism. To what extent this approach limits risk description is discussed later.

Let us describe a basic example of frequency risk model. The statistical units are indexed by $i = 1, \ldots, n$, and the dependent variable is a sequence of claims numbers. We denote it as

$$Y_i = (Y_{i,t})_{t=1,\ldots,T_i};\ Y_{i,t} \sim P(\lambda_{i,t}),\ \lambda_{i,t} = \exp(x_{i,t}\beta).$$

A duration $d_{i,t}$ of risk exposure must be included in the parameter of the Poisson distribution if these durations are not constant on the sample. In the *a priori* rating model, the variables $Y_{i,t}$ are independent and this property also holds in the fixed effects model

$$Y_{i,t} \sim P(\lambda_{i,t}\, u_i).$$

The reference value of the time-independent fixed effect u_i is one. If $u_i > 1$, the individual i is riskier than the average of its peers with respect to the regression components.

The random effects model (where the fixed effect u_i is the outcome of U_i) can be defined parametrically, with an explicit distribution for U_i. The distribution of $Y_{i,t}$ is defined by an expectation with respect to U_i, i.e.

$$P[Y_{i,t} = n] = E\left[P_{\lambda_{i,t}\, U_i}(n)\right] = E\left[\exp(-\lambda_{i,t}\, U_i) \times \frac{\lambda_{i,t}^n\, U_i^n}{n!}\right].$$

With Gamma distributions, $(U_i \sim \gamma(a,a) :\ E(U_i) = 1,\ V(U_i) = 1/a)$, the distributions of the risk variables are negative binomial. Extensions of the negative binomial model to panel data are given in Hausman et al. (1984).

A semiparametric specification stems from the equation

$$E(U_i) = 1 \Rightarrow E(Y_{i,t}) = \lambda_{i,t};\ V(Y_{i,t}) = \lambda_{i,t} + \left(\lambda_{i,t}^2 \times V(U_i)\right) \tag{1}$$

in the random effects model. It appears that the variance σ^2 of the random effect is the natural parameter of the mixing distribution in a semiparametric approach.

The prediction $\widehat{u_i}^{T_i+1}$ of the fixed effect u_i with a linear credibility approach stems from a linear probabilistic regression of U_i with respect to the $Y_{i,t}$ ($t = 1, \ldots, T_i$) in the random effects model. The solution is

$$\widehat{u_i}^{T_i+1} = \frac{1 + \left(\widehat{\sigma^2} \times \sum_{t=1}^{T_i} y_{i,t}\right)}{1 + \left(\widehat{\sigma^2} \times \sum_{t=1}^{T} \widehat{\lambda_{i,t}}\right)}, \tag{2}$$

where a consistent estimation of the variance of the random effect is obtained from (1) as

$$\widehat{\sigma^2} = \frac{\sum_{i,t}\left[(y_{i,t} - \widehat{\lambda_{i,t}})^2 - \widehat{\lambda_{i,t}}\right]}{\sum_{i,t} \widehat{\lambda_{i,t}}^2}. \tag{3}$$

The predictor of equation (2) is also obtained with an expected value principle in a Poisson model with Gamma random effects (Dionne and Vanasse, 1989). The semiparametric estimator of the variance is unconstrained and is positive only if there is overdispersion on the data (i.e. if the residual variance is greater than the empirical mean).[7] A consistent estimation strategy of the parameters of a random effects models is detailed in the next section in a semiparametric framework. This strategy exploits two results that are obtained in this example.

- First, the expectation of the risk variable in the random effects model does not depend on σ^2. As a consequence, the estimation of β in the Poisson model is consistent in the model with random effects.
- Second, equation (3) provides an estimator of σ^2 that depends on $\widehat{\beta}$. This is due to a separability property in the specification of the variance of the risk variable in the random effects model.

The prediction of the fixed effect u_i obtained from a posterior expectation in the negative binomial model is the same as that obtained with the linear credibility approach. This predictor can be written as a weighted average of $1 = E(U_i)$, and of the ratio $\sum_t y_{i,t} / \sum_t \widehat{\lambda_{i,t}}$, which summarizes the individual history and which can be seen as an estimator of the fixed effect u_i. The weight given to this ratio is the credibility

$$cred_i = \frac{\widehat{\sigma^2} \times \sum_t \widehat{\lambda_{i,t}}}{1 + \left(\widehat{\sigma^2} \times \sum_t \widehat{\lambda_{i,t}}\right)}, \tag{4}$$

[7] We have $\sum_{i,t} \widehat{\lambda_{i,t}} = \sum_{i,t} y_{i,t}$ from the orthogonality between the residuals and the intercept.

which ranges in [0, 1] and increases with risk exposure (represented by the cumulated frequency premium) and the estimated variance $\widehat{\sigma^2}$, which represents the weight of unobserved heterogeneity. From the weighted average definition of the predictor, the credibility is the discount on the frequency premium (the "bonus") if no claims are reported.

The experience rated premium for the next period is $\widehat{\lambda}_{i,T_i+1} \times \widehat{u_i}^{T_i+1}$. The predictor $\widehat{u_i}^{T_i+1}$ summarizes the individual history and can be interpreted as a "bonus-malus" coefficient. From equation (2), the estimated variance of the random effect is close to the relative increase in premium after a claim (the "malus") if risk exposure is close to 0.

The linear shape of the predictor in this example can be challenged. Prediction with a posterior expectation would not be linear in the number of past claims if the mixing distribution was not of the Gamma type. We might want to obtain other shapes as the exponential one in the French bonus-malus system.

The parametric and semiparametric approaches of risk prediction both rely on restrictions. The mixing distribution family is constrained in the parametric approach, whereas the shape of the predictor is constrained in the semiparametric setting. Discarding these restrictions is possible with a non parametric analysis of the mixing distribution. Such approaches are feasible, but they can be applied only with high frequency data, which is not the case in non-life insurance. To see this, consider the moment result on Poisson distributions

$$Y \sim P(\lambda) \Rightarrow E\left[Y \times (Y-1) \ldots \times (Y-k+1)\right] = \lambda^k \ \forall k \in \mathbb{N}^*. \tag{5}$$

If Y follows a mixture of a $P(\lambda u)$ distributions, where u is the outcome of a random effect U, we have that

$$E(U) = 1 \Rightarrow E(U^k) = \frac{E\left[Y \times (Y-1) \ldots \times (Y-k+1)\right]}{\lambda^k = \left[E(Y)\right]^k}. \tag{6}$$

Then the mixing distribution can be identified from a sequence of moments of increasing order (i.e., going from a semiparametric to a non parametric approach through a representation of the mixing distribution by moments of increasing order).[8] However equation (6) suggests that the accuracy of the estimation of a high order moment of the random effect is weak if the frequency risk $E(Y)$ is low. This is the case in non-life insurance, and explains why experience rating models restrict to parametric and semiparametric approaches.

3.1 Statistical models on count data of the (a,b,k) type

This article deals mostly with frequency risk models, and we present a distribution family on count data that encompasses the usual ones. This distribution family on

[8]See Zhang (1990) for an approximation of the Fourier transform of the mixing distribution.

$\mathbb{N}$ is defined from $(a,\, b,\, k)$, (with $0 < a < 1$, $b > 0$, and $k \in \mathbb{N}$) in the following way (Klugman, Panjer, and Willmot, 2008):

- If p_n is the probability related to $n \in \mathbb{N}$, the $(p_n)_{n<k}$ are defined without any constraints other than their belonging to the simplex of probabilities.
- The tail distribution is defined from the recurrence relation

$$p_n = p_{n-1} \times \left(a + \frac{b}{n}\right),\; n > k. \tag{7}$$

This equation allows to denote the ratio $(\sum_{n>k} p_n)/p_k$ as $M(a, b, k)$. Then the tail distribution is defined from

$$p_k = \frac{1 - \sum_{n<k} p_n}{1 + M(a, b, k)}$$

and from equation (7).

Let us recover usual distribution families on count data as distributions of the $(a,\, b,\, k)$ type.

- A Poisson distribution $P(\lambda)$ is obviously of the $(0, \lambda, 0)$ type.
- A "zero inflated" distribution linked with a variable $B \times N$, where $B \sim B(1, p)$ and $N \sim P(\lambda)$ are independent variables (see Boucher, Denuit, and Guillén (2009) for applications to insurance rating) is of the type $(0, \lambda, 1)$, with $p_0 = \exp(-\lambda) + [(1 - p) \times (1 - \exp(-\lambda))] \geq \exp(-\lambda)$.
- Let us consider a negative binomial distribution, obtained as a mixture of $P(\lambda u)$ distributions, where u is the outcome of U, $U \sim \gamma(a, a)$. It is easily seen that this distribution is of the type $\left(\frac{\lambda}{\lambda+a}, \frac{\lambda\times(a-1)}{\lambda+a}, 0\right)$. Hence, all the distributions of the $(a,\, b,\, 0)$ type are either of the Poisson or of the negative binomial type.

Distributions of the $(a,\, b,\, k)$ type, with $k > 1$ can be considered if the frequency is not too low.

4 Estimation approaches for random effects models

4.1 The Generalized Estimating Equations

Statistical models are designed depending on the nature of the dependent variable. For instance, a binary distribution is defined by its expectation, and the model deals with the link between this expectation (and the related probability) and regression components. Going from the most constrained distribution in terms of support (the binary distributions) to the less constrained (variables that range on the whole real line) allows to disconnect completely the expectation and moments of higher order, including the variance. Between these two polar cases, non-life insurance models first deal with count variables, where events are insurance claims. The claim frequency per year is usually far less than one, which constrains the design of statistical models as mentioned in the preceding section.

The Generalized Estimating Equations approach (Zeger et al., 1986) proposes an estimation strategy from the a second order specification of the moments of a dependent risk variable, that can be applied for frequency risk and linear models. Let i be a statistical unit in a sample of size n, and let Y_i be a risk variable ranging in $\mathbb{R}^{d_i}$. The statistical unit may include time series, strata, multiple equations related to different guarantees or to a frequency-cost specification, etc. The expectation and the variance of Y_i are denoted as

$$E(Y_i \,|\, x_i,\, \beta),\; V(Y_i \,|\; x_i,\, \beta,\, \alpha).$$

The parameters $\beta,\, \alpha$ ($\beta \in \mathbb{R}^{k_1},\; \alpha \in \mathbb{R}^{k_2}$) of the model are included hierarchically, and the specific parameters of the mixing distribution represented by α do not influence the expectation of the risk variable.[9] These specific parameters are usually second-order moments of random effects. These random effects are plugged additively in the expectation of Y_i for linear models and multiplicatively for frequency risk models. The independence of $E(Y_i)$ with respect to α is obtained from obvious constraints on the expectation of the random effects in the additive and multiplicative setting. These specifications also yield separability properties which allow to estimate α from β and the observations, using cross-section moment equations. Let us consider a statistic $M(y_i, x_i, \alpha, \beta)$ such as

$$\alpha,\; M(y_i, x_i, \alpha, \beta) \in \mathbb{R}^{k_2};\; E\left[M(Y_i, x_i, \alpha, \beta) \,|\; x_i,\, \beta,\, \alpha\right] \equiv 0;$$
$$\frac{\partial}{\partial \alpha} M(Y_i, x_i, \alpha, \beta) \text{ is invertible.} \tag{8}$$

We have for instance: $M(y_i, x_i, \alpha, \beta) = \sum_t \left[(y_{i,t} - \lambda_{i,t})^2 - \lambda_{i,t}\right] - \sigma^2 \sum_t \lambda_{i,t}^2$ for the basic example developed in the preceding section, where $\alpha = \sigma^2$ is the variance of a scalar random effect. We suppose that

$$\sum_{i=1}^{n} M(y_i, x_i, \alpha, \beta) = 0 \Leftrightarrow \exists!\; \alpha,\; \alpha = \widehat{\alpha}(\beta;\, y_1, \ldots, y_n;\, x_1, \ldots, x_n).$$

[9] The independence of the random effects distribution with respect to the regression components can be challenged. This issue is dicussed by Boucher, Denuit (2006) and by Bolancé, Guillén, Pinquet (2008).

This condition is linked to the invertibility condition given in (8), and the solution α does not necessarily belongs to the parameter set, as is the case for the example if there is underdispersion.

The algorithm $\widehat{\beta}^m$, $\widehat{\alpha}^m \to \widehat{\beta}^{m+1}$, $\widehat{\alpha}^{m+1}$ is then the following: first, the variances-covariances matrices of risk units

$$\widehat{V}_i^m = V(Y_i \mid x_i, \widehat{\beta}^m, \widehat{\alpha}^m)$$

are derived from the current estimations of the parameters. Then the estimations at the next step are obtained as follows:

$$\widehat{\beta}^{m+1} = \arg\min_{\beta} \sum_i ||y_i - E(Y_i \,|\, x_i, \beta)||^2_{[\widehat{V}_i^m]^{-1}} = \arg\min_{\beta} f(\beta, \widehat{\beta}^m, \widehat{\alpha}^m); \qquad (9)$$

$$\widehat{\alpha}^{m+1} = \widehat{\alpha}(\widehat{\beta}^{m+1};\ y_1, \ldots, y_n;\ x_1, \ldots, x_n).$$

The algorithm can be initialized at step $m = 0$ with $\widehat{\alpha}^0 = 0$, which corresponds to no unobserved heterogeneity, and with $\widehat{\beta}^0 = \arg\min_{\beta} f(\beta, \beta, 0)$, with the notations of equation (9).

This estimated approach is semiparametric and unconstrained with respect to the parameters of the mixing distribution.[10] An estimation obtained outside the parameter domain is a failure of the model which corresponds to an estimation obtained at the boundary of the parameter set with a constrained estimation approach. In the example studied in this article, a negative estimation for the variance σ^2 of the random effect corresponds to a residual underdispersion on the data. A maximum likelihood estimation of a parametric mixture of Poisson distributions would lead to a null estimated variance. Indeed, the numerator of the ratio which defines the unconstrained estimator of the variance given in (3) is equal to twice the Lagrangian of the log-likelihood with respect to σ^2 at the frontier of the parameter set. Then underdispersion leads to a local maximum of the likelihood, which actually is global. When the mixing distribution family is more intricate, a constrained estimation obtained at the boundary of the parameter set may indicate feasible submodels more clearly than an unconstrained approach.

4.2 Other estimating approaches

Let us consider first a parametric setting. The likelihood of a random effects model is an expectation, which does not have a closed form in most cases. The likelihood can be then approximated, and two types of computation can be investigated.

• Numerical integration of the likelihood. If the likelihood is viewed as a parameter, the approximation is a biased and deterministic estimator. See Davis and

[10] The parameter set for α is usually a convex cone in $\mathbb{R}^{k_2}$.

Rabinowitz (1984) for methods of numerical integration using Gaussian quadrature rules, and Lillard (1993) for empirical results.

• Monte-Carlo methods interpret the likelihood as the expectation of a function of a distribution-free variable. An average derived from independent draws of this variable for each individual leads to a simulation-based estimator. The likelihood is then approximated by a random and unbiased variable. Owing to the concavity of the logarithm, the estimator of the log-likelihood has a negative bias. The asymptotic properties of these estimators are given by Gouriéroux and Monfort (1991). Consistency is obtained if the number of simulations converges towards infinity with the size of the sample.

We come back to a semiparametric setting. In the Generalized Estimating Equations approach presented in Section 4.1, the first and second order moments of the dependent variable have implicitly a closed form in the random effects model. However this assumption does not hold in most cases for binary variables. Suppose that these moments are approximated by simulations. If the simulation errors are independent across observations and sufficiently regular with respect to the parameters, the simulation-based estimators can be consistent even if the number of draws are fixed for each individual. Consistency is obtained if a linearity property allows the simulation errors to be averaged out over the sample. A proof of these properties and applications to discrete response models are found in Mac Fadden (1989).

5 The nature of the dynamics on non-life insurance data

Random effect models reflect the observed dynamics on non-life insurance data, as estimated risks usually decrease with time and increase with claims. This time-event property fits the "bonus-malus" logic of risk prediction based on random effects models. Two points will be developed further.

- The first point is the analysis of the data dynamics. The observed dynamics on risks reflect both revelation and modification effects of the individual histories. The revelation effect of unobserved heterogeneity is not intrinsic, as it is defined with respect to the observable information. The individual histories modify the risks levels due to incentive effects (the *homo œconomicus* reacts to the financial implications of his behaviour), but also to psychological effects that influence risk perception and tastes. These effects usually counteract the revelation effects, but this is not always true and will be discussed in the next section.

- The second point is the identifiability issue of the two components of the dynamics. The main motivation is to analyze the incentive effects of insurance rating. This point is also analyzed by Chiappori and Salanié (2012).

5.1 Incentives effects of non-life insurance rating schemes

The incentive properties of an insurance rating scheme are obtained from the minimization of the lifetime disutility of future premiums. The incentive level is related to the increase in the future premiums after a claim. From the exponential structure of the French bonus-malus system, a claim at fault (which triggers the "malus") increases the incentives to safe driving (see Abbring, Chiappori, Pinquet, 2003). The risk level decreases after a claim, which counteracts the revelation effect of unobserved heterogeneity. However an opposite effect could be obtained if the potential penalties did not increase after a claim (i.e. if the premium was not a convex function of the number of past claims, for a given risk exposure).

The time effects of incentives are at the opposite of the event effect, and the relative weights depend on the equilibrium of the rating scheme. Let us consider the French bonus-malus system. A 25% malus balances a 5% bonus if the annual frequency of claims is close to 1/6. The frequency of claims at fault is actually equal to 6%, and the French bonus-malus system is downwards biased, as is the case for most of the experience rating schemes (see Lemaire, 1995). Drivers cluster at the lowest levels of the bonus-malus scale when their seniority increases, and are subject on average to decreasing incentives. This means that the time effect of incentives outweigh the event effect in this context. It is worth mentioning that the result also depends on the frequency risk of the driver.

The time effect of incentives can reinforce the revelation effect if the reward for a claimless history consists in cancelling the claim record after a given duration. This feature exits in the French bonus-malus system for drivers with a bad claim history. If their bonus-malus coefficient is greater than one (that of a beginner), they are considered as beginners after a two year claimless history.[11] An informal argument to explain this result is that the date of claim removal does not vary with time, and that safe driving effort increases as this date comes nearer. A more formal argument is that the incentive level increases with the difference between the lifetime disutility of premiums in the state reached after a claim, and the disutility in the current state. The time counter is reset to zero after a claim, and the disutility after a claim is constant. As the current disutility decreases with time, the difference increases with time and so does the incentive level. Then risk decreases with time

[11] The same logic is applied in many point-record driving licenses (where events are traffic violations which are associated to demerit points, and where the driving license is suspended once the cumulated demerit points reach a given threshold). In France and in many European countries, all the demerit point are removed after a given period of violation-free driving. In the US and in Canada, point removal is performed on each traffic offense once a given seniority is reached. The incentive properties of point-record mechanisms are studied by Bourgeon and Picard (2007), and by Dionne, Pinquet et al. (2011).

as for the revelation effect of unobserved heterogeneity. Hence incentive effects do not always counteract revelation effects in non-life insurance.

5.2 Identifiability issues in the analysis of non-life insurance data dynamics

Early statistical literature did not grasp the identification issue raised by the interpretation of individual histories. Discussing a paper written by Neyman (1939), Feller (1943) mentions the two interpretations of the negative binomial model with revelation and modification stories. These two interpretations of data dynamics are also termed as heterogeneity and state dependence. Feller remarks that this twofold interpretation is not understood by most of statisticians, including Neyman.[12] Feller concludes to the impossibility of identifying the nature of the dynamics of longitudinal data. At the end of his article, he suggests that a duration-event analysis could help improve identification.

This article was taken seriously by Neyman, who wrote an article with Bates a decade later (Bates and Neyman, 1952) proposing an elimination strategy of unobserved heterogeneity for a point process of the Poisson type. They restrict their analysis to individuals with a single event observed on a given interval. The date distribution of this event is uniform, and a Kolmogorov-Smirnov test of fit to a uniform distribution allows to integrate out unobserved heterogeneity in the test for the absence of state dependence, according to Neyman. Many papers in econometric literature (see Chiappori and Salanié (2012), for a survey) stem from this contribution.

Bates and Neyman's conclusion is however overoptimistic. Indeed, a mixture of Poisson processes can be applied to real individuals and not to a class of individuals sharing the same available information. In that case, the history modifies the individual distributions instead of revealing them, although the null assumption tested for by Bates and Neyman is fulfilled. For instance, a mixture of a Poisson process with a parameter λ and a $\gamma(a,a)$ mixing distribution is associated to a Markov process with integer values, where the only positive transition intensities are those from n to $n+1$ ($n \in \mathbb{N}$), and equal to

$$\lambda_n(t) = \lambda \frac{a+n}{a+\lambda t}$$

at date t. We obtain a Pólya process, with negative binomial marginal distributions. The date distribution of a unique event in a given interval is also uniform, as we show now. Let us consider an interval $[0,T]$, and N_t the number of events between

[12] Neyman was far from being a beginner when he wrote this paper. He already had published his results on optimal tests with Egon Pearson.

0 et t. We denote $\Lambda_n(t) = \int_0^t \lambda_n(u)\, du$. We have that

$$P[N_T = 1] = \int_0^T \exp(-\Lambda_0(t))\lambda_0(t)\exp(\Lambda_1(t) - \Lambda_1(T))\, dt,$$

where t is the date of the unique event. The density of this date is equal to $\lambda_0(t) \times \exp(\Lambda_1(t) - \Lambda_0(t))$, up to a multiplicative constant. The log-derivative of the density is equal to $\left(\lambda_0^{'}/\lambda_0\right) + \lambda_1 - \lambda_0$. The null assumption tested for by Bates and Neyman reflects an equilibrium between the time and event components of the data dynamics, i.e.

$$\frac{\lambda_0^{'}}{\lambda_0} \text{ (time)}; \;\; \lambda_1 - \lambda_0 \text{ (event)}. \tag{10}$$

In the Pólya process, we have $\lambda_0^{'}/\lambda_0 < 0$, and $\lambda_1 - \lambda_0 > 0$. But opposite signs can be observed for these components if they are related to incentives derived from a convex rating structure, as discussed in the preceding section. The time-event psychological effects can also be represented by equation (10). The "availability effect" is associated to an increasing link between time and risk, and to a decreasing event-risk link. Results are at the opposite for the "gambler's fallacy" effect. As a conclusion, what is eliminated by the Bates-Neyman test is unobserved heterogeneity applied to balanced time-event effects on real individuals.[13]

6 Examples of frequency and pure premium risk models

6.1 Multiple equations and stratified samples

Different types of claims can be used in the prediction of non-life insurance risks, as for instance claims at fault and not at fault, accidents and traffic violations (in a framework where traffic violations are used for experience rating and not accidents in a no-fault environment). These different types of claims can be nested (e.g. accidents with bodily injury among accidents of all type in automobile insurance: see section 6.3), or overlap partially or not. In a situation where event types (e.g. type A and type B) overlap partially, a random effects model should be applied to non overlapping events (e.g. $A - B$, $B - A$ and $A \cap B$). Mixing distributions can be estimated in a semiparametric framework (Pinquet, 1998) or with parametric specifications (Frees and Valdez, 2008). For small risk exposures, it can be shown that the predictive ability of a given type of event on another type in a frequency risk model is proportional to the product of the frequency risk and of the squared

[13]The test proposed by Abbring, Chiappori, and Pinquet (2003) eliminates unobserved heterogeneity in some unbalanced time-event frameworks.

covariance of the random effects related to each type and applied multiplicatively to the frequency.

Stratified samples are for examples fleets of vehicles (Desjardins et al. (2001), Angers et al. (2006)), whether owned by companies or households. The history of a contract should have a greater ability to predict the risk level of this contract than that of the other contracts in the same stratum. The relative efficiencies are obtained from the comparison between the variance of a random effect at the stratum level and the residual variance at the individual level.

6.2 Allowance for the age of claims in experience rating

Real-world experience rating schemes in property-liability insurance mostly depend on numbers of events, which are usually claims at fault. Only in a few publications (Gerber and Jones, 1975; Sundt, 1988; Bolancé, Guillén and Pinquet, 2003) do frequency risk models take into account the age of events. These contributions use the intuition that the predictive ability for a period of the policyholder's history should decrease with age. If a stationary specification is retained for time-varying random effects in a Poisson model, the estimated autocorrelation coefficients should be decreasing. This shape is indeed usually obtained from non-life insurance data.

With time-independent random effects, total credibility converges to one as frequency risk exposure increases (see equation (4), and remember that credibility is the no-claim discount related to a claimless history). This result does not hold any more with dynamic random effects. Limit credibility can be much less than one, a result in accordance with real-world rating structures.

A bonus-malus system designed from a model with dynamic random effects and a decreasing autocorrelation function will behave in the following way. For a policyholder with a faultless history, the no-claim discounts induced by a claimless year are smaller after a few years than those obtained from the usual credibility model, but they are more important if claims were reported recently. The explanation is the same in both cases. The credibility granted to a given period of the past decreases rapidly as time goes by, due to the increase of risk exposure but mostly to the diminution of the autocorrelation coefficients. Notice that economic analysis also suggests that optimal insurance contracts with moral hazard should penalize recent claims more than older ones (Henriet and Rochet, 1986).

Dynamics on longitudinal count data can also be obtained from endogeneous approaches. The integer autoregressive model of order one (or INAR(1) model) writes as follows for a single time series

$$N_t = I_t + \sum_{j=1}^{N_{t-1}} B_{j,t}.$$

The number of events at the current period, is the sum of two independent variables. The first variable is a number of events without link to those occurred in the past, and represents an innovation. The second variable is a sum of Bernoulli variables indexed by the events occurred at the preceding period, and provides a causality relationship between events. If I_t and N_{t-1} are Poisson variables and if the Bernoulli variables are i.i.d., N_t is also a Poisson variable. Parameters are retained in order to ensure the stationarity of N_t.

With the INAR model, the predictive ability of past events decreases with seniority, which is in accordance with real-life data. However the autocorrelation structure is similar to that of a linear process, and this feature does not fit the data in non-life insurance. Let us consider the covariances between a time series of count variables. In a Poisson model with dynamic random effects, these covariances are obtained from the autocorrelation coefficient applied to the overdispersion of the count variable. With the INAR specification, the autocorrelation coefficient is applied to the total variance of the count variable, and data speak in favor of the preceding formulation in non-life insurance.[14] Considering mixtures of INAR processes can alleviate this shortcoming (see Gouriéroux and Jasiak, 2004).

6.3 Allowance for the severity of claims in experience rating

Multi equation models can be used to allow for the severity of claims involving third party liability, from the dichotomy between claims with or without bodily injury (see Picard (1976), Lemaire (1995)). The number of claims with bodily injury follows a binomial distribution, indexed by the number of claims and by a probability which follows a beta distribution in the random effects model. Nesting this random effect in a negative binomial model yields a linear predictor based on the number of claims of both types.

For guarantees related to property damage or theft, a cost equation on claims can be considered. Gamma or log-normal distributions provide a good fit to cost data without thick tails.[15] A two equation model with Poisson distributions for the number of claims and log-normal distributions for their cost admits closed form estimators for the second order moments of bivariate random effects (Pinquet (1997)). The correlation between the random effects related to the number and cost equations appears to be very low for the sample investigated in the aforementioned article. Because of this low correlation, the bonus-malus coefficients related to pure premium are close to the product of the coefficients for frequency and expected cost per claim.

[14] Also, the prediction derived from the INAR(1) model is derived from the number of events restricted to the last period. This is an unpleasant property if events are claims, as all the claims in the past have some predictive ability.

[15] Log-normal distributions have however thicker tails than the gamma, as they are of the subexponential type.

In a recent publication, Frees and Valdez (2008) propose a three equation model corresponding to the frequency, type and cost of claims. The first equation is a random effects Poisson regression model, the second is a multinomial logit model, and the cost component is a Burr XII long-tailed distribution. A t-copula function is used to specify the joint multivariate distribution of the cost of claims arising from these various claims types.

References

[1] Abbring, J., P.A. Chiappori and J. Pinquet (2003) "Moral Hazard and Dynamic Insurance Data," *Journal of the European Economic Association* **1**, 767-820.

[2] Angers J.F., D. Desjardins, G. Dionne and F. Guertin (2006) "Vehicle and Fleet Random Effects in a Model of Insurance Rating for Fleets of Vehicles," *ASTIN Bulletin* **36**, 25-77.

[3] Bates, G.E., and J. Neyman (1952) "Contributions to the Theory of Accident Proneness II: True or False Contagion," *University of California Publications in Statistics* **1,** 255-275.

[4] Bolancé C., M. Guillén and J. Pinquet (2003) "Time-Varying Credibility for Frequency Risk models: Estimation and Tests for Autoregressive Specifications on the Random Effects," *Insurance: Mathematics and Economics* **33**, 273-282.

[5] Bolancé C., M. Guillén and J. Pinquet (2008) "On the Link between Credibility and Frequency Premium," *Insurance: Mathematics and Economics* **43**, 209-213.

[6] Boucher J.-P., M. Denuit (2006) "Fixed Versus Random Effects Models in Poisson Regression Models for Claim Counts: Case Study with Motor Insurance," *ASTIN Bulletin* **36**, 285-301.

[7] Boucher, J.-P., M. Denuit and M. Guillén (2009) "Number of Accidents or Number of Claims? An Approach with Zero-Inflated Poisson Models for Panel Data," *Journal of Risk and Insurance,* **76**, 821-846.

[8] Bourgeon, J.-M. and P. Picard (2007) "Point-Record Driving License and Road Safety: An Economic Approach," *Journal of Public Economics* **91**, 235-258.

[9] Bühlmann, H. (1967) "Experience Rating and Credibility," *ASTIN Bulletin* **4**, 199-207.

[10] Bühlmann, H. and A. Gisler (2005) *A Course in Credibility Theory and its Applications*. Springer Verlag, Universitext.

[11] Chiappori, P.A. and G. Salanié (2012) "Asymmetric Information in Insurance Markets: Empirical Assessments," *Handbook of Insurance,* Springer Verlag (Editor: Georges Dionne).

[12] Davis, P., and P. Rabinowitz (1984) *Methods of Numerical Integration.* Academic Press, New York.

[13] Denuit, M., X. Marechal, S. Pitrebois and J.-F. Wahlin (2007) *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Scales.* Wiley, New-York.

[14] Desjardins D., G. Dionne, and J. Pinquet (2001) "Experience Rating Schemes for Fleets of Vehicles," *ASTIN Bulletin* **31**, 81-106.

[15] Dionne, G., and C. Vanasse (1989) "A Generalization of Automobile Insurance Rating Models: the Negative Binomial Distribution with a Regression Component," *ASTIN Bulletin* **19**, 199-212.

[16] Dionne, G., J. Pinquet, C. Vanasse and M. Maurice (2011) "Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data," *The Review of Economics and Statistics* **93**, 218-227.

[17] Ehrlich, I. and G.S. Becker (1972) "Market Insurance, Self-Insurance, and Self-Protection," *Journal of Political Economy* **80,** 623-648.

[18] Feller, W. (1943) "On a General Class of "Contagious" Distributions," *The Annals of Mathematical Statistics* **14**, 389-400.

[19] Frees, E.W. and E. Valdez (2008) "Hierarchical Insurance Claims Modeling," *Journal of the American Statistical Association* **103,** 1457-1469.

[20] Gerber, H., and D. Jones (1975) "Credibility Formulas of the Updating Type," *Transactions of the Society of Actuaries Vol.* **27***,* 31-52.

[21] Gouriéroux, C., and A. Monfort (1991) "Simulation Based Inference in Models with Heterogeneity," *Annales d'Economie et de Statistiques* **20-21**, 69-107.

[22] Gouriéroux, C. and J. Jasiak (2004) "Heterogeneous INAR(1) model with application to car insurance," *Insurance: Mathematics and Economics* **34**, 177-192.

[23] Hausman, J.A., B.H. Hall, and Z. Griliches (1984) "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* **52**, 909-938.

[24] Hendel, I., and A. Lizzeri (2003) "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *The Quarterly Journal of Economics* **118**, 299-327.

[25] Henriet, D., et J.C. Rochet (1986) "La Logique des Systèmes Bonus-Malus en Assurance Automobile," *Annales d'Economie et de Statistiques* **1,** 133-152.

[26] Klugman, S., H. Panjer and S. Willmot (2008) *Loss Models: from Data to Decisions.* Wiley, New-York.

[27] Kunreuther, H. and M.V. Pauly (1985) "Market Equilibrium with Private Knowledge: An Insurance Example," *Journal of Public Economics* **26**, 269-288. Reprinted in *Foundations of Insurance Economics,* Kluwer Academic Publishers (editors: G. Dionne and S. Harrington).

[28] Lemaire, J. (1995) *Bonus-Malus Systems in Automobile Insurance.* Huebner International Series on Risk, Insurance and Economic Security.

[29] Lillard, L. (1993) "Simultaneous Equations for Hazards (Marriage Duration and Fertility Timing)," *Journal of Econometrics* **56**, 189-217.

[30] Mac Fadden, D. (1989) "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica* **57**, 995-1026.

[31] Nelder, J.A., and R.W.M. Wedderburn (1972) "Generalized Linear Models," *Journal of the Royal Statistical Society A* **135**, 370-384.

[32] Neyman, J. (1939) "On a New Class of "Contagious" Distributions, Applicable in Entomology and Bacteriology," *The Annals of Mathematical Statistics* **10,** 35-57.

[33] Nini, G., and P. Kofman (2011) " Do Insurance Companies Exploit an Informational Monopoly? Empirical Evidence from Auto Insurance," working paper, The Wharton School.

[34] Picard, P. (1976) "Généralisation de l'Etude sur la Survenance des Sinistres en Assurance Automobile," *Bulletin Trimestriel de l'Institut des Actuaires Français,* 204-267.

[35] Pinquet, J. (1997) "Allowance for Cost of Claims in Bonus-Malus Systems," *ASTIN Bulletin* **27**, No. 1, 33-57.

[36] Pinquet, J. (1998) "Designing Optimal Bonus-Malus Systems from Different Types of Claims," *ASTIN Bulletin* **28**, No. 2, 205-220.

[37] Pinquet, J. (2000) "Experience Rating through Heterogeneous Models," *Handbook of Insurance* 459-500, Kluwer Academic Publishers. Huebner International Series on Risk, Insurance and Economic Security (Editor: Georges Dionne).

[38] Sundt, B. (1981) "Credibility Estimators with Geometric Weights," *Insurance: Mathematics and Economics* **7**, 113-122.

[39] Taylor, G. (1986) "Underwriting Strategy in a Competitive Insurance Environment," *Insurance: Mathematics and Economics* **5***,* 59-77.

[40] Tversky, A. and D. Kahneman (1973) "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* **5**, 207-232.

[41] Tversky, A. and D. Kahneman (1974) "Judgment under Uncertainty: Heuristics and Biases," *Science* **185**, 1124-1131.

[42] Zeger, S.L., Liang, K.Y., and P.S. Albert (1986) "Models for Longitudinal Data: A Generalized Estimating Equations Approach" *Biometrics* **44**, 1049-1060.

[43] Zhang, C.H. (1990) "Fourier Methods for Estimating Densities and Distributions," *The Annals of Statistics* **18,** 806-831.