The Geneva Papers on Risk and Insurance Theory, 24: 119–137 (1999)

The Econometrics of Risk Classification in Insurance

C. GOURIEROUX
CREST and CEPREMAP

Abstract

We present in this article some questions related to risk classification. These are discussed depending on the information used—either data on conditional characteristics or also including data on claim histories or on endogenous insurance demand by the agents.

Key words: risk classification, adverse selection, moral hazard, Poisson-gamma model, bonus-malus

1. Introduction

The significance of the econometric (or statistical) insurance analysis can be appraised by considering the "production process" of an insurance company (see figure 1). An insurer is a financial intermediary that offers and sells a set of contracts, which may be partly reinsured. This activity entails some negative cash flows including the reinsurance premiums and the residual costs of the insured claims, and some positive cash flows corresponding to the premiums payed by the insured agents. These negative and positive cash flows do not have symmetric patterns, and temporarily positive or negative balances may arise. These divergences can be smoothed out by adequate hedging strategies on financial markets. In this situation, the company has to determine jointly the selection of contracts, the reinsurance policies, the pricing strategies, and the updating of the hedging portfolios. Clearly the right decisions can only be taken if the firm has accurate information on its own costs, which originate mainly from either direct or indirect[1] costs of the claims. The main challenge for the statistician is to predict the occurrence of claims, their severities, and their costs.

This prediction problem leads to the following issues:

1. Even if we are ultimately interested in the prediction of an aggregate, i.e., the final outcome of the firm, the analysis of claims has to be carried out contract by contract and not directly on the portfolio of contracts. This is the so-called *risk classification*. Indeed, the insured agents involve various risks, which may be taken into account to eventually differentiate the prices, the reinsurance policies, or the hedging strategies.[2] Moreover, to examine the evolution of the risk contained in a portfolio, it is crucial to distinguish between the dynamics of the risk in each class and the modification of the partition structure of the insured agents. Indeed, the second factor may be partly controlled by the firm contract selection or by a differentiated pricing scheme.
2. While the properties of the basic financial assets are well summarized by the expected returns and volatilities, i.e., by the first- and second-order (conditional) moments of the

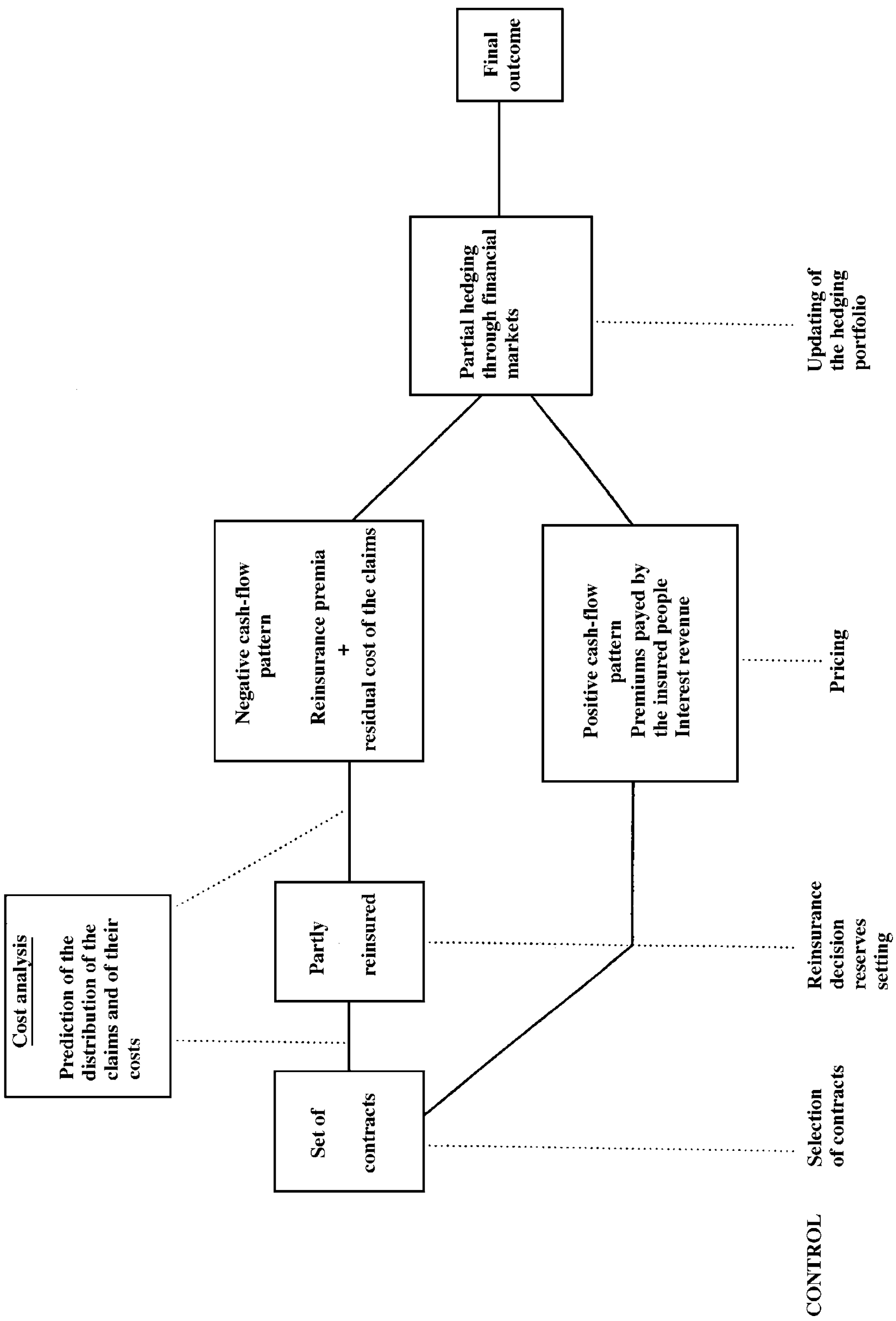
Cost analysis
Prediction of the distribution of the claims and of their costs
Set of contracts
Partly reinsured
Negative cash-flow pattern
Reinsurance premia
+
residual cost of the claims
Positive cash-flow pattern
Premiums payed by the insured people
Interest revenue
Partial hedging through financial markets
Final outcome
CONTROL
Selection of contracts
Reinsurance decision reserves setting
Pricing
Updating of the hedging portfolio

Figure 1. The “production process” of an insurance company.

returns, it is necessary to consider in the insurance contracts the whole distribution of the claim cost. There are two reasons: first, these distributions are far from Gaussian because they admit a point mass at zero (no claim, zero cost); and second, in complex contracts including deductibles or in partly reinsured contracts, the cost is a nonlinear transformation of the cost of the corresponding basic contract without deductible and reinsurance. In some sense, these complex contracts are derivatives, and, typically in financial theory, the analysis of derivatives requires the knowledge of the whole distribution of the underlying asset cash flow.

In this article, we are essentially interested in the question of risk classification and in the distributional assumptions that may be done. In Section 2, we first consider the static case, where the contracts concern a given period. We define carefully the notion of class of risks and discuss the criteria that may be followed to select a classification. The case of a sequence of contracts is analysed in Section 3. We explain how the individual histories may be taken into account to improve the risk classification under moral hazard and adverse selection when more information becomes available. From the example of car insurance, we show how the introduction of unobservable heterogeneity factors may be used to derive simple updating formulas for the classification and the pure premium. Finally, in Section 4, we consider the additional information that may be introduced under adverse selection by proposing different contracts to the agent and considering which one has been selected.

2. Static risk classification

The basic models used in portfolio risk analysis are adapted to a static framework. A typical setup consists of a population $\mathcal{P}$ of individuals, indexed by i, $i = 1, \dots, n$. The claims of interest are defined as either a scalar or vector[3] variable, say Y_i, that has to be predicted by the insurance company. The prediction is performed on the basis of some observed individual variables x_i that allows classification of the individuals. In practice, the covariates x_i are used to define a partition of the whole population:

$$\mathcal{P} = \bigcup_{k=1}^{K} \mathcal{P}_k,$$

where $\mathcal{P}_k = \{i : x_i \in A_k\}$, $k = 1, \dots, K$, and A_k, $k = 1, \dots, K$ is a partition of the set of possible values of x_i. The partition is defined by crossing some qualitative characteristics, such as age, gender, occupation, and so on.

Clearly, this procedure involves some degree of arbitrariness. It has to be emphasized that the selection of the partition is essential for a feasible, accurate, and robust assessment of the portfolio risk. Below, we discuss general requirements that need to be satisfied for this purpose (see Gourieroux [1999] for a more detailed discussion).

2.1. Homogenous classes of risks

The feasibility requirement concerns the possibility of aggregating the individual risk in order to derive the portfolio risk and of estimating the unknown parameters of the model by averaging on subpopulations. This requirement determines the constraints imposed on the partition structure for the statistical inference.

Definition 1: *A partition is composed of homogenous classes of risks if and only if*

(i) *the conditional distribution of $Y_1, \ldots, Y_n$ given $X_1, \ldots, X_n$ coincides with the conditional distribution of $Y_1, \ldots, Y_n$ given $Z_1, \ldots, Z_n$, where Z_i indicates the class to which individual i belongs*;
(ii) *$Y_1, \ldots, Y_n$ are independent conditionally on $Z_1, \ldots, Z_n$; and*
(iii) *the conditional distribution of Y_i only depends on the class of i.*

The first condition means that the information content of the partition is equivalent to that of the initial personal characteristics x. The two remaining conditions are introduced to simplify the aggregation procedure.

For illustration, let us consider a standard mean-variance framework, where the distribution of a scalar risk Y is summarized by its mean and variance, denoted μ and σ^2, respectively. The above definition clearly implies heterogeneity, since μ_k, σ_k^2 will generally depend on the class $\mathcal{P}_k$. Therefore the individual risks may be displayed in a mean-variance graph for comparison (see figure 2).

Note that the term *homogenous* is used above with the following statistical interpretation: in a given class, the individual risks are independent, with identical distributions. We emphasize that this does not mean that the individuals of a given class have identical claims. Indeed, in the usual classifications, we may encounter some classes with small variances, but also some groups for which the observed variables provide poor information. To give

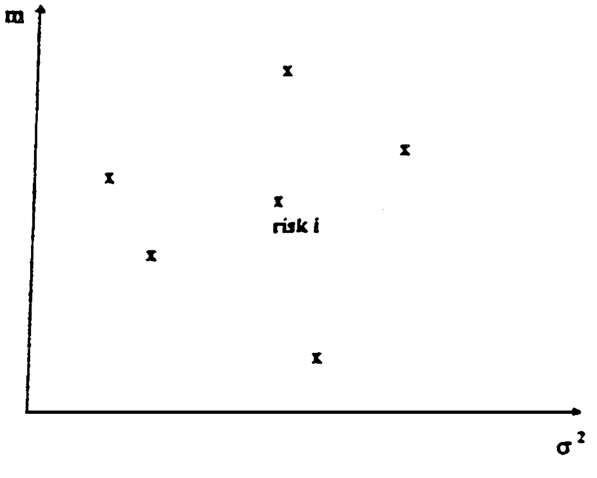

Figure 2. Mean-variance representation of individual risks.

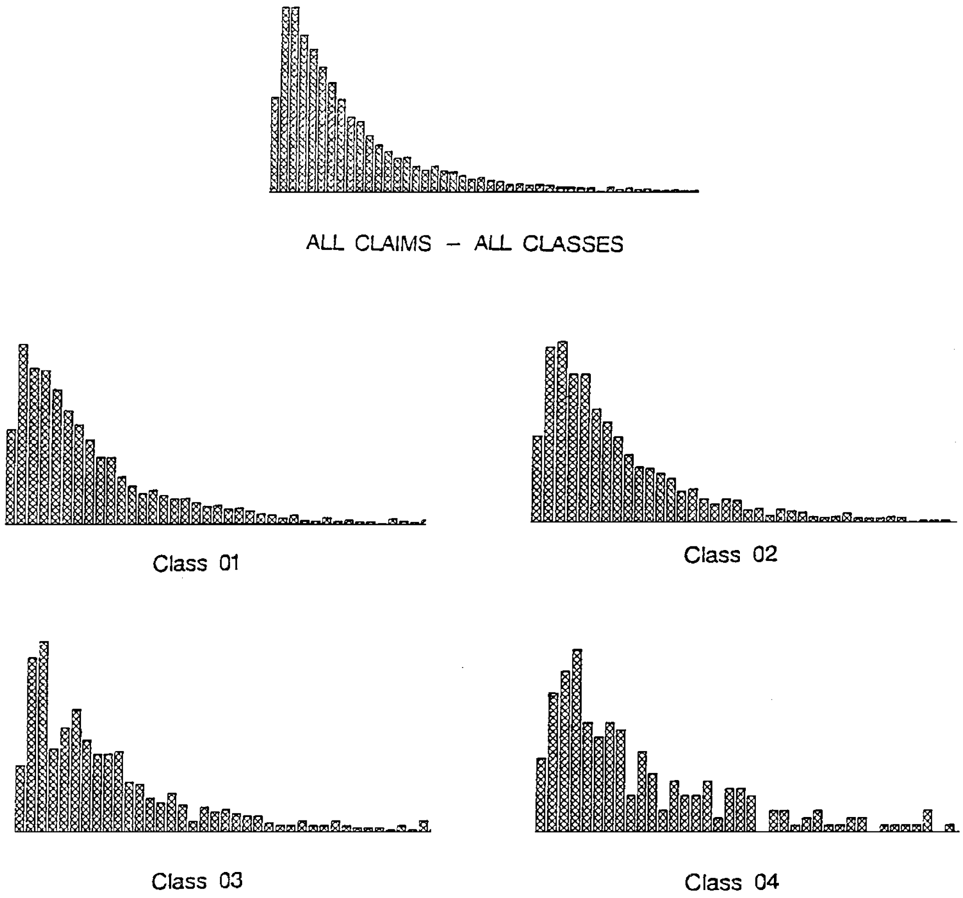

Figure 3. Cost distributions in car insurance.

some idea of this heterogeneity effect between and within classes, we provide in figure 3 the distribution of the annual cost of claims for car insurance. The data concern a population of insured people from Quebec. The first graph provides the average cost in each group (a class defined by the company) and the other ones the distribution within each retained group. It may be noted that the (conditional) distributions present some regular form, with only one mode and some asymmetry. A model may be based on either log-normal, log-logistic (see Beirlandt et al. [1991]), or gamma distribution.

Let us now consider a portfolio containing n_k contracts of type $k, k = 1, \dots, K$. The global risk $Y = \sum_{i\in\mathcal{P}} Y_i = \sum_{k=1}^{K}(\sum_{i\in\mathcal{P}_k} Y_i)$ is summarized by its mean, $\mu = \sum_{k=1}^{K} n_k m_k$, and its variance, $\sigma^2 = \sum_{k=1}^{K} n_k \sigma_k^2$. Hence, under this partition, standard aggregation formulas apply.

At this point, it becomes clear why the risk analysis has to be primarily performed at the individual rather than the portfolio level. Indeed, the difference between the characteristics of two portfolios at two different dates t and τ, namely,

$$\mu_t = \sum_{k=1}^{K} n_{kt} m_{kt}, \quad \sigma_t^2 = \sum_{k=1}^{K} n_{kt} \sigma_{kt}^2,$$
$$\mu_\tau = \sum_{k=1}^{K} n_{k\tau} m_{k\tau}, \quad \sigma_\tau^2 = \sum_{k=1}^{K} n_{k\tau} \sigma_{k\tau}^2,$$

may be due to any modification of the individual risks, reflected by the variation of m_k and σ_k^2, or any changes in the portfolio composition, i.e., n_k, $k = 1, \ldots, K$.

2.2. *Precision and robustness*

Common sense suggests that the thinner the partition, the more accurate is the evaluation of the portfolio risk. However, in our setup this intuition turns out to be misleading for two reasons.

By increasing the partition, we may lose some properties required for the classes of risk, especially the conditional independence between the Y_i variables; moreover, as the size of the class decreases, the estimation of the unknown parameters μ_k, σ_k^2 becomes less precise.

In summary, the selection of an adequate partition involves a tradeoff between small within variances, robustness, and feasibility.

2.3. *The distributional specifications*

As pointed out in the introduction, several insurance contracts may be considered as derivatives, due to the deductible or reinsurance effects. For this reason, the distribution in each class of risk has to be fully specified. It is usually selected from a standard family of distributions, depending on the interpretation of the Y variable—for example, the occurrence of an accident, number of claims in a given period, duration before the first claim, severity, cost, etc.

When these families are parameterized, they have to depend at least on two parameters in order to cover all the positions in the mean-variance space. For example, let us consider the standard Poisson model for count variables, where $Y_i \sim \mathcal{P}(\lambda_k)$, if i belongs to class k.

This distribution is such that the mean and variance coincide. Therefore, in the mean-variance space the only admissible positions of this class are on the 45° line (see figure 4). In order to avoid such a restrictive specification, several authors (see, e.g., Gourieroux, Monfort, and Trognon [1984], Cameron and Trivedi [1986], [1999], Boyer, Dionne, and Vanasse [1992], Lemaire [1995]) have proposed to enlarge this class by introducing an additional error term. It is assumed that the term retains some heterogeneity on the λ parameter for the individuals belonging to the same class k, i.e.,

$$Y_i / \mu_i \sim \mathcal{P}[\lambda_k \mu_i],$$

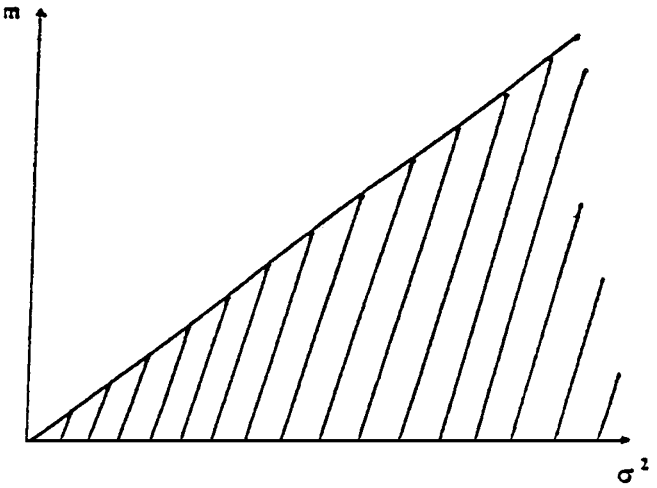

Figure 4. The Poisson and negative binomial models in the mean-variance space.

where μ_i is some independent unobservable heterogeneity factor. If μ_i follows a gamma distribution $\gamma(A_k, A_k)$, which implies $E\mu_i = 1$, then Y_i may be shown to follow a negative binomial distribution, with two parameters depending on λ_k and A_k and in a one-to-one relationship with μ_k, σ_k^2, where $\sigma_k^2 \geq \mu_k$. Note that the use of a negative binomial distribution should be the basis for a test of price discrimination in insurance. Indeed, if this specification is not rejected, the heterogeneity factor should be priced and a price discrimination should follow. The negative binomial model has been implemented in a number of empirical studies (see, e.g., Richaudeau [1997], Pinquet [1997a]).

3. The risk dynamics

Agents are generally insured for several consecutive periods, either by the same company or by different ones. In any case, the risk has to be considered in a dynamic framework, since it is probably time varying and since the available information on individual behaviors increases with time (see, e.g., Pinquet [1997b]). In particular, we may know the claims submitted by the agent in the past.

There are three questions to be considered.

1. What is the informational content of the individual claim histories with respect to the static individual characteristics?
2. How should risk evaluation be updated—for instance, every year?
3. What is the effect of the prediction horizon?

The natural answer to the first question is that the role of individual history increases with time. Individual history may even reveal the effect of some unobserved individual characteristics on the risk and may allow prediction of the values of these risks. Sometimes this characteristic may be controlled by the individual. A typical example is vehicle speed in car insurance. This is strongly related to the notion of the *effort variable* usually introduced to discuss moral hazard phenomena.

The updating of the predicted risks may be performed along different lines. We may focus on some structural specification of the risk distribution, estimate it, and then derive the updating formulas either analytically or numerically. Alternatively, we may directly propose some updating formulas with simple interpretations that will only be optimal for a specific underlying structural model. The modification of our knowledge of the individual risk will generally imply a modification of the risk classification and of the insurance premiums, the so-called *bonus-malus* in car insurance.

Finally, we have to study how the risk classification depends on the prediction horizon.

In the two next subsections, we illustrate these points by using some selected models from car insurance analysis.

3.1. A model with unobserved heterogeneity

Let us consider the problem of car insurance. We denote by $Z_{i,t}$ the number of claims submitted by individual i in period t, and by $Y_{i,t}$ the corresponding total cost.[4] To simplify the presentation, we assume that the only observed characteristics are individual variables x_i and past histories of claims, summarized by the lagged values of $Z_{i,t}$, and $Y_{i,t}$. We also introduce some unobserved time-invariant individual variables μ_i, α_i, which influence the number of claims and the cost, respectively. Therefore, we have a double heterogeneity: the first is directly observed (by x_i), and the second (μ_i, α_i) is hidden and partly revealed by the observed driving.

We now introduce a set of distributional assumptions on the different unobserved variables $\{(Z_{i,t}, Y_{i,t}), t = 1, \ldots, T, \mu_i, \alpha_i\}, i = 1, \ldots, n$.

A.1. *The pairs* $(Z_{i,t}, Y_{i,t})$ $t = 1, \ldots, T, i = 1, \ldots, n$ *are independent conditionally on* x_i, $\mu_i, \alpha_i, i = 1, \ldots, n$.

A.2. *The distribution of the number of claims* $Z_{i,t}$ *conditional on* x_i, μ_i, α_i, $i = 1, \ldots, n$ *is a Poisson distribution*:

$$Z_{it}/x_i, \mu_i, \alpha_i \sim \mathcal{P}[\mu_i \exp x_i b], t = 1, \ldots, T.$$

A.3. *The distribution of the total cost* $Y_{i,t}$ *conditional on* $Z_{it}, x_i, \mu_i, \alpha_i$, $i = 1, \ldots, n$ *is a gamma distribution*:

$$Y_{i,t}/Z_{i,t}, x_i, \mu_i, \alpha_i \sim \gamma(\nu Z_{i,t}, \alpha_i \exp(-x_i d)), t = 1, \ldots, T.$$

A.4. *The unobserved characteristics* μ_i, α_i $i = 1, \ldots, n$ *are independent, conditionally on the observed characteristics. Their distributions are*

$$\mu_i/x_i \sim \gamma(A, A),$$
$$\alpha_i/x_i \sim \gamma(C + 1, C).$$

In these assumptions, b, d, A, C, and ν are parameters to be estimated: b and d give the effects of the explanatory variables on the number of claims and the cost by claim, respectively; A and C measure the magnitude of heterogeneity. The distributions are selected according to the type of variable, namely, Poisson for the count variable and gamma for the variable with positive values. Moreover, the distributions are conjugate, which ensures a simple analytical derivation of the marginal and conditional distributions and of the updating formulas (see the Appendix).

Finally, note that the heterogeneity parameters μ_i and α_i are defined up to a multiplicative factor whenever a constant is introduced among the explanatory variables x_i. Therefore, we may impose restrictions on their gamma distributions. The normalizations have been chosen such that $E(\mu_i) = E(1/\alpha_i) = 1$.

We may deduce from the previous assumptions the p.d.f. of the observed endogenous variables $Z_{i,t}$, and $Y_{i,t}$ $t = 1, \dots, T$ given the observed exogenous variables x_i only. By multiplying across the individuals, we derive the likelihood function for estimation of the unknown parameters. The p.d.f. is (see the Appendix)

$$\begin{aligned} & f(z_{i,t}, y_{i,t}, t = 1, \dots, T/x_i) \\ & = \prod_{t=1}^{T} \left\{ \frac{\exp(z_{it} x_i b)}{z_{it}!} \frac{y_{it}^{\nu z_{it} - 1}}{\Gamma(\nu z_{it})} \right\} \\ & \frac{\Gamma\left(\sum_{t=1}^{T} z_{it} + A\right)}{\Gamma(A)} \frac{A^A}{(T \exp x_i b + A)^{\sum_{t=1}^{T} z_{it} + A}} \\ & \frac{\Gamma\left(\nu \sum_{t=1}^{T} z_{it} + C + 1\right)}{\Gamma(C+1)} \frac{C^{C+1}}{[\exp(-x_i d) \sum_{t=1}^{T} y_{it} + C]^{\nu \sum_{t=1}^{T} z_{it} + C + 1}}, \end{aligned}$$

where the small letters z, y are introduced for the observed values of the variables.

Let us now consider the updating formulas. We have two risk summaries, namely, the number of claims and the cost by claim. This implies that the updating formulas will probably require bivariate recursive equations. We essentially consider below the prediction of the endogenous variables $Z_{i,T}$ and $Y_{i,T}$, and not the evaluation of the prediction accuracy. Let us define

$$\begin{aligned} \hat{Z}_{i,T+1} &= E[Z_{i,T+1}/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i], \\ \hat{Y}_{i,T+1} &= E[Y_{i,T+1}/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i], \end{aligned}$$

where $\underline{Z_{i,T}} = (Z_{i,T}, Z_{i,T-1}, \dots)$ and $\underline{Y_{i,T}} = (Y_{i,T}, Y_{i,T-1}, \dots)$ are the observed individual histories.

The modification of the predicted total cost between T and $T + 1$ is performed by the double recursion

$$\begin{aligned} \hat{Z}_{i,T+1} - \hat{Z}_{i,T} &= \frac{\exp x_i b}{T \exp x_i b + A} (z_{i,T} - \hat{Z}_{i,T}), \\ \frac{\hat{Y}_{i,T+1}}{\hat{Z}_{i,T+1}} - \frac{\hat{Y}_{i,T}}{\hat{Z}_{i,T}} &= \frac{z_{i,T}}{\frac{(T-1)\exp x_i b + A}{\exp x_i b} \hat{Z}_{i,T-1} + \frac{C}{\nu} - A} \left(\frac{y_{i,T}}{z_{i,T}} - \frac{\hat{Y}_{i,T}}{\hat{Z}_{i,T}} \right). \end{aligned}$$

This is an adaptive scheme in which the predicted number of claims and the predicted cost by claim are updated on the basis of the most recent prediction errors. The adjustment speeds are not fixed. For instance, for the number of claims, the prediction error is given by $\frac{\exp x_i b}{T \exp x_i b + A}$, which clearly varies not only with the individual but also with the time of the last observation. Hence the new prediction error matters relatively less in the update for large T.

The predicted $Y_{i,T}$ generally depends simultaneously on x_i, the individual histories, and the parameters of the model. In the limiting case $T \to \infty$, we can see that

$$\lim_{T \to \infty} \hat{Y}_{i,T} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} y_{it}.$$

This prediction is determined only by the lagged observations on claims submitted by the individual and does not involve data on the individuals belonging to the same class.

From a statistical point of view, a heterogeneity factor may be considered as an unknown individual parameter, for which we have introduced a distribution to model our lack of knowledge. Hence the approach may be considered to be of the Bayesian type and is strongly related with the so-called *credibility theory* (see Buhlman [1967] and the list of references therein). This approach has been applied above in the context of bivariate nonlinear risks.

3.2. A model with stochastic conditional heterogeneity

The aforementioned prediction property in the limit arises as the consequence of the heterogeneity specified by the model. Recall that the two unobservable characteristics μ_i, λ_i were assumed to be time independent. Under the effort-variables interpretation, the individual does not adapt his/her effort level to the time-varying information. This assumption is quite unrealistic and probably not fulfilled. Moreover, it yields the result showed above, i.e., the possibility of entirely individualized insurance when T is large.

The insurance premium corresponding to this model is a rough measure of the cost of the insurance contract—a measure that does not accommodate the value of risk associated with the time variation of the individual's effort.

We propose below a method allowing an extension of the model in order to overcome this difficulty. This approach was previously used to extend ARCH models into stochastic variance models (see Ghysels, Harvey, and Renault [1997] for a complete discussion of this question) or to define stochastic volatility duration models (Ghysels, Gourieroux, and Jasiak [1997]). It has been described in a general framework by Gourieroux and Jasiak [1999]. Since the simplest dynamic models applicable to error terms are the vector autoregressive models (VARs), we first transform the basic heterogeneity factors into Gaussian variables. Let us denote by Φ the c.d.f. of the standard normal distribution and by $H(\cdot\,; A, C)$ the cumulative function of the $\gamma(A, C)$ distribution. If ε follows the $\gamma(A, C)$ distribution, the transformed variable $\Phi^{-1}[H(\varepsilon; A, C)]$ follows a standard normal distribution.

The previous set of assumptions is modified by introducing path-dependent heterogeneity factors $\mu_{i,t}, \alpha_{i,t}$ such that

$$\begin{aligned} Z_{i,t}/x_i, \mu_{i,t}, \alpha_{i,t} &\sim \mathcal{P}[\mu_{i,t} \exp x_i b], \\ Y_{i,t}/Z_{i,t}, x_i, \mu_{i,t}, \alpha_{i,t} &\sim \gamma[\nu Z_{i,t}, \alpha_{it} \exp(-x_i d)]. \end{aligned}$$

In the next step, these heterogeneity factors are normalized and written as

$$\begin{aligned} \mu^*_{i,t} &= \Phi^{-1}[H(\mu_{i,t}; A, A)], \\ \alpha^*_{i,t} &= \Phi^{-1}[H(\alpha_{i,t}; C+1, C)]. \end{aligned}$$

A VAR structure (for instance, of order 1) can be directly imposed on these transformed factors:

$$\begin{pmatrix} \mu^*_{i,t} \\ \alpha^*_{i,t} \end{pmatrix} = \psi \begin{pmatrix} \mu^*_{i,t-1} \\ \alpha^*_{i,t-1} \end{pmatrix} + u_{i,t},$$

where ψ is a 2 × 2 matrix and $(u_{i,t})$ is Gaussian white noise. This approach allows pricing of the instantaneous volatilities on the heterogeneity factors along with their persistence degrees. Moreover, this specification requires modified summary statistics of individual histories. Because of the autoregressive pattern followed by the heterogeneity factors, the summary statistics will differentiate the past claims by associating larger weights with the most recent ones. This is a way to introduce geometric weights in a nonlinear framework (see Gerber and Jones [1973, 1975] and Sundt [1981] for this question in the context of credibility theory).

3.3. A priori and a posteriori classifications

At the initial date, i.e., without information on past claims, the risk classification may only be based on the individual variables x_i. When more information becomes available, the optimal predictions can be constructed from both x_i and the summary statistics of the past claims, i.e., either

$$\tilde{x}_{i,T} = (\hat{Z}_{i,T}, \hat{Y}_{i,T}),$$

or

$$\tilde{x}_{i,T} = \left(\frac{1}{T} \sum_{t=1}^{T} z_{i,t}, \frac{1}{T} \sum_{t=1}^{T} y_{i,t} \right)$$

(see the Appendix). In this case, the initial (or a priori) risk classification, based on x_i only, can be distinguished from the a posteriori classification determined jointly by $(x_i, \tilde{x}_{i,T})$.

In such a framework, the information on individual past claims cannot be summarized by a single bonus-malus coefficient, since $\tilde{x}_{i,T}$ has two components, and the past behaviors of the drivers cannot be totally ordered. The components can be ranked only when we consider a specific insurance contract. Let us consider a policy for insuring the total cost $Y_{i,T+1}$ with a deductible level L. The corresponding pure premium is, say,

$$E[(Y_{i,t+1} - L)^+/x_i, \tilde{x}_{i,T}] = g(x_i, \tilde{x}_{i,T}; L).$$

For each deductible level, the previous formula explains how to weight the two components of $\tilde{x}_{i,T}$. Some drivers may appear less risky than others for a small deductible level and simultaneously more risky for a large deductible level.

The usefulness of considering two components for the risk and introducing the two associated heterogeneity factors is now clear. Let us consider the standard model for car insurance, which is based on the count variable only (or which assumes, alternatively, that the cost by accident is independent of the occurrence and of the individual characteristics). The risk characteristics are summarized by the moments conditional on the available information $I_{i,T}$, namely,

$$E[Z_{i,T+1}/I_{i,T}],\ V(Z_{i,T+1}/I_{i,T}),$$

which are in a linear relationship (see the Appendix)

$$V[Z_{i,T+1}/I_{i,T}] = E(Z_{i,T+1}/I_{i,T}) \left\{1 + \frac{\exp x_i b}{T \exp x_i b + A}\right\}.$$

Therefore, the locations in the mean-variance space are very constrained (see figure 5). We may also note that these locations tend to the 45° line when T tends to infinity. At the limit, the heterogeneity factor μ_i is entirely known, and we get the Poisson model.

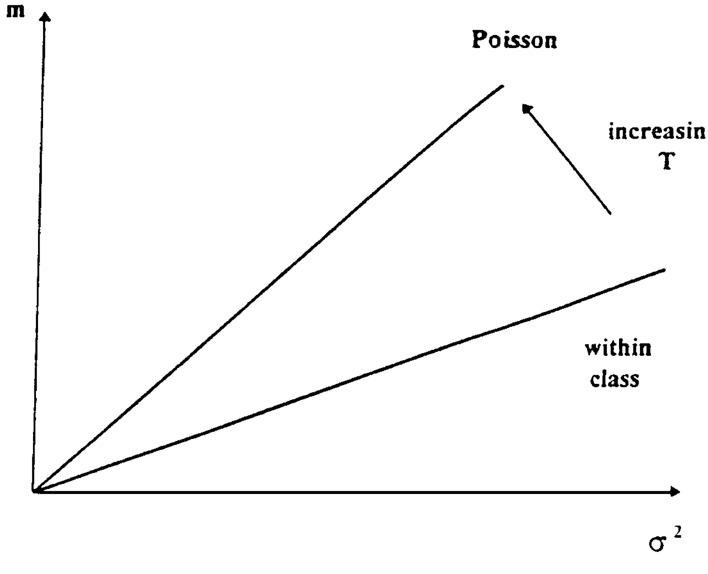

Figure 5. Within risk for the count model.

3.4. *Self-predictive classification*

The previous discussion essentially concerns the prediction of the individual claims one period ahead. In practice, it is important to analyze also the properties of the portfolio over larger horizons. Indeed, the predicted risk may be small one period ahead but may increase considerably over two periods.

Let us denote by H the prediction horizon. To predict $Z_{i,T+H}$ and $Y_{i,T+H}$ using the information available at time T, we compute

$$\begin{aligned}\hat{Z}_{i,T,T+H} &= E[Z_{i,T+H}/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i],\\ \hat{Y}_{i,T,T+H} &= E[Y_{i,T+H}/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i],\end{aligned}$$

in our illustration. These predictions depend on the past history by means of summary statistics, $\tilde{x}^H_{i,T}$, which may be functions of the horizon H. Whenever this is true, the risk classifications also become H dependent.

It may be interesting to identify some models and variables such that the risk classifications are valid for all horizons. This can be done by imposing constraints on the classification variables. We discuss this issue in the context of the car insurance example. From the iterated expectations theorem, we have

$$\begin{aligned}\hat{Z}_{i,T,T+H} &= E[Z_{i,T+H}/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i]\\ &= E\{E[Z_{i,T+H}/\underline{Z_{i,T+H-1}}, \underline{Y_{i,T+H-1}}, x_i]/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i\}\\ &= E\{E[Z_{i,T+H}/\tilde{x}_{i,T+H-1}, x_i]/\underline{Z_{i,T}}, \underline{Y_{i,T}}, x_i\}.\end{aligned}$$

Therefore, we have to predict some function of the summary statistics at date $T + H - 1$ from the available information and to verify if the prediction depends on the information through $\tilde{x}_{i,T}$, x_i only. We deduce that the classification remains valid over all the horizons if it holds both for evaluating the risk at horizon 1 and for predicting the future dynamics of the individual among the classes of risks (the condition of self-predictive classification).

Whenever the self-predictive classification condition is fulfilled, the risk at horizon H may be predicted by first predicting the location of the individual among the classes of risk at date $T + H - 1$ and then predicting the risk at horizon 1 conditionally on this predicted location.

The condition may be fulfilled for some particular models. Let us consider the previous Poisson–gamma model. We know that the conditional distribution for horizon 1 is such that

$$l(y_{i,T+1}, z_{i,T+1}/I_{i,T}) = l(y_{i,T+1}, z_{i,T+1}/\tilde{x}_{i,T}, x_i).$$

We deduce that

$$\begin{aligned}&l(y_{i,T+2}, z_{i,T+2}, y_{i,T+1}, z_{i,T+1}/I_{i,T})\\ &\quad = l(y_{i,T+2}, z_{i,T+2}/I_{i,T+1})l(y_{i,T+1}, z_{i,T+1}/I_{i,T})\\ &\quad = l(y_{i,T+2}, z_{i,T+2}/\tilde{x}_{i,T+1}, x_i)l(y_{i,T+1}, z_{i,T+1}/\tilde{x}_{i,T}, x_i).\end{aligned}$$

Since

$$\begin{aligned}\tilde{x}_{i,T+1} &= \left(\sum_{t=1}^{T+1} z_{i,t}, \sum_{t=1}^{T+1} y_{i,t}\right),\\ &= \tilde{x}_{i,T} + (z_{i,T+1}, y_{i,T+1}),\end{aligned}$$

we get

$$\begin{aligned}&l(y_{i,T+2}, z_{i,T+2}, y_{i,T+1}, z_{i,T+1}/I_T)\\ &\quad = l(y_{i,T+2}, z_{i,T+2}/\tilde{x}_{i,T} + (z_{i,T+1}, y_{i,T+1}), x_i) l(y_{i,T+1}, z_{i,T+1}/\tilde{x}_{i,T}, x_i).\end{aligned}$$

By integrating with respect to $z_{i,T+1}, y_{i,T+1}$ conditionally on $\tilde{x}_{iT}, x_i$, we conclude that

$$l(y_{i,T+2}, z_{i,T+2}/I_{i,T}) = l(y_{i,T+2}/\tilde{x}_{i,T}, x_i).$$

The same summary statistics $\tilde{x}_{i,T}$ may be used for horizons 1 and 2 in the Poisson–gamma model.

4. Self-selectivity effects

An individual may select one particular insurance contract among the different ones proposed by the insurance company; for instance, he or she may choose the deductible level. An important question concerns the information on his or her risk, which may be deduced from his or her choice.

4.1. *Optimal behavior*

For illustration, let us consider the example of car insurance with time-independent heterogeneity factors. Two types of contracts are proposed: the first without deductible for a premium $p_{i,t}$ in period t and the second with a deductible level L for a premium $q_{i,t}$. If the individual is risk neutral, performs rational expectations, and has complete knowledge of his or her characteristics x_i and heterogeneity factors μ_i, α_i, he or she will select the policy without deductible at time t if and only if

$$\begin{aligned}&p_{i,t} - E[Y_{i,t+1}/\underline{Y_{i,t}}, \underline{Z_{i,t}}, x_i, \mu_i, \alpha_i]\\ &\quad < q_{i,t} - E[(Y_{i,t+1} - L)^+/\underline{Y_{i,t}}, \underline{Z_{i,t}}, x_i, \mu_i, \alpha_i].\end{aligned} \tag{1}$$

Let us denote by $\xi_{i,t}$ the selection variable

$$\begin{aligned}\xi_{i,t} = {}&1, \text{ if no deductible is chosen at period } t,\\ &0, \text{ otherwise.}\end{aligned}$$

We may write

$$\xi_{i,t} = \begin{cases} 1, & \text{if } g(\underline{Y_{i,t}}, \underline{Z_{i,t}}, x_i, \mu_i, \alpha_i, p_{it}, q_{it}; \theta) > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where g is the function deduced from (1) and θ is the parameter of the initial model. Under this optimal behavior, the observation of the selected alternative brings additional information on the values of the unobservable heterogeneity factors.

4.2. Practical implementation

However, it is not clear that such asymmetric information between the insured agent and the insurance company exists or that the agent has rational behavior, a full knowledge of $\mu_i, \alpha_i, \ldots,$ etc. Hence, in practice, we have to verify if the observed choice (possibly including past choices) brings additional information. For this purpose, we may consider the conditional distribution of $Z_{i,t}$ and $Y_{i,t}$, given the past histories $\underline{Z_{i,t-1}}$, $\underline{Y_{i,t-1}}$, the current and past choices $\underline{\xi_{i,t}}$, and the observable individual characteristics x_i. If the choices are noninformative, the conditional distribution will be the same as in Section 3. Otherwise, the current and past choices will be significant. Hence we may extend the basic model by introducing $\xi_{i,t}$—for instance, as an additional explanatory variable, taking into account some possible cross-effects with $x_i, \tilde{x}_{i,t-1}$—and look for the significance of the associated coefficients. The conclusion depends on the risk classification initially introduced. The available data show that in a sufficiently sharp class of risks, almost all individuals select the same deductible level, which implies no additional information. This result is confirmed by recent econometric studies by Chiappori-Salanie [1996] and Dionne, Gourieroux, and Vanasse [1997], where it is interpreted as the absence of residual adverse selection.[5]

4.3. Dynamic analysis of the heterogeneity factors

The dynamic relation between the individual risk and demand may also be analyzed along the following lines. Let us consider the model with stochastic conditional heterogeneity introduced in Section 3.2. From the observable $Z_{i,t}, Y_{i,t}, t = 1, \ldots, T$, and the model, we may deduce some predicted values for the heterogeneity factors $\hat{\mu}_{i,t}, \hat{\alpha}_{i,t}$ $t = 1, \ldots, T$ (say). Then we get a bivariate time series, whose evolution may be compared to the evolution of the demand performed by the agent $\xi_{i,t}$ $t = 1, \ldots, T$ (say) and of the premiums $p_{i,t}, q_{i,t}, t = 1, \ldots, T$. A complete causality analysis between these series may allow the detection of some moral hazard phenomena.

5. Concluding remarks

We have presented in this article some questions related to risk classification. These have been discussed depending on the information used—either data on conditional characteristics or also including data on claim histories or on endogenous insurance demand by the agent.

From the example of car insurance, we have introduced several extensions, such as the distinction of two risk components corresponding to the occurrence and the cost of the claim, respectively, the two heterogeneity factors to capture the claim history by means of at least two summary statistics, the possibility of a stochastic evolution of these underlying factors, at the basis of moral hazard testing procedure. This example has been followed to illustrate a progressive econometric analysis, and the same kind of approach may be applied to more complex risks and contracts (see, e.g., Gourieroux and Scaillet [1997] for an application to unemployment insurance on mortgages).

Appendix: Derivation of the marginal and conditional distributions

Distributions

Let us assume that

$$\begin{aligned} Z_{i,t}/x_i, \mu_i, \alpha_i &\sim \mathcal{P}[\mu_i \exp x_i b], \\ Y_{i,t}/Z_{i,t}, x_i, \mu_i \alpha_i &\sim \gamma[\nu Z_{i,t}, \alpha_i \exp(-x_i d)], \\ \mu_i \sim \gamma(A, A), \alpha_i &\sim \gamma(C+1, C). \end{aligned}$$

The joint p.d.f. of $Z_{i,1}, \dots, Z_{iT}, Y_{i,1}, \dots, Y_{i,T}, \mu_i, \alpha_i$ is

$$\exp[-\mu(A + T \exp x_i b)]\, \mu^{\sum_{t=1}^T z_{it} + A - 1} \frac{(\exp x_i b)^{\sum_{t=1}^T z_{it}} A^A}{\prod_{t=1}^T (z_{it}!)\Gamma(A)}$$
$$\exp\left[-\alpha\left(\exp x_i d \sum_{t=1}^T y_{it} + C\right)\right] \alpha^{\nu \sum_{t=1}^T z_{it} + C} (\exp - x_i d)^{\sum_{t=1}^T z_{it}}$$
$$\frac{C^{C+1}}{\prod_{t=1}^T \Gamma(\nu z_{it}) \Gamma(C+1)} \prod_{t=1}^T [y_{it}^{\nu z_{it} - 1}].$$

We deduce the property below and the p.d.f. of the observable variables given in Section 3.

Property: *The heterogeneity factors μ_i and α_i are independent conditionally on $I_{i,T} = (Z_{i,t}, Y_{i,t})$ $t = 1, \dots, T$, with distributions*

$$\mu_i / I_{i,T} \sim \gamma\left(\sum_{t=1}^T z_{i,t} + A, T \exp x_i b + A\right),$$

$$\alpha_i / I_{i,T} \sim \gamma\left(\nu \sum_{t=1}^T z_{i,t} + C + 1, \exp(-x_i d) \sum_{t=1}^T y_{it} + C\right).$$

Prediction formulas

This property may be directly used to compute the predictions of $Z_{i,T+1}$, $Y_{i,T+1}$ at date T. We have

$$\begin{aligned}
&E(Z_{i,T+1}/I_{i,T})\\
&\quad = E[E(Z_{i,T+1}/I_{i,T}, \mu_i, \alpha_i)/I_{i,T}] \quad \text{(from the theorem of iterated expectations)}\\
&\quad = E[\mu_i \exp(x_i b)/I_{i,T}]\\
&\quad = \exp(x_i b)\frac{\sum_{t=1}^{T} z_{it} + A}{T \exp x_i b + A};\\
&E(Y_{i,T+1}/I_{i,T})\\
&\quad = E[E(Y_{i,T+1}/I_{i,T}, \mu_i, \alpha_i)/I_{i,T}]\\
&\quad = E\left(\nu \exp x_i(b+d)\frac{\mu_i}{\alpha_i}\Big/ I_{i,T}\right)\\
&\quad = \nu \exp x_i(b+d) E(\mu_i/I_{i,T}) E\left(\frac{1}{\alpha_i}\Big/ I_{i,T}\right)\\
&\quad = \nu \exp x_i(b+d)\frac{\sum_{t=1}^{T} z_{it} + A}{T \exp x_i b + A}\,\frac{\exp(-x_i d)\sum_{t=1}^{T} y_{it} + C}{\nu \sum_{t=1}^{T} z_{it} + C}.
\end{aligned}$$

The updating formulas follow.

Conditional variances

We can compute the conditional variances similarly. For instance, let us consider the count variable $Z_{i,T+1}$. We get

$$\begin{aligned}
&V(Z_{i,T+1}/I_{i,T})\\
&\quad = E[V(Z_{i,T+1}/\mu_i, I_{i,T})/I_{i,T}] + V[E(Z_{i,T+1}/\mu_i, I_{i,T})/I_{i,T}]\\
&\quad = E[\mu_i \exp x_i b/I_{i,T}] + V[\mu_i \exp x_i b/I_{i,T}]\\
&\quad = \exp(x_i b) E(\mu_i/I_{i,T}) + \exp(2x_i b) V[\mu_i/I_{i,T}]\\
&\quad = \exp(x_i b)\frac{\sum_{t=1}^{T} z_{i,t} + A}{T \exp x_i b + A}\left\{1 + \frac{\exp x_i b}{T \exp x_i b + A}\right\}.
\end{aligned}$$

We then deduce the linear relationship between the two first conditional moments of the count variable:

$$V[Z_{i,T+1}/I_{i,T}] = E[Z_{i,T+1}/I_{i,T}]\left\{1 + \frac{\exp x_i b}{T \exp x_i b + A}\right\}.$$

Acknowledgment

We have benefited from helpful comments by G. Dionne, J. Jasiak, A. Monfort, and C. Vanasse.

Notes

1. Cost of reinsurance, control, and soforth.
2. The usual financial models applied to insurance assume a dynamic model for the value of the portfolio with a small number of factors, generally one (see, e.g., Chang, Chung, and Kimsky [1989]). Such practice neglects the time-varying risk heterogeneity.
3. Especially for multirisk analysis.
4. The risk data are assumed to be summarized by the two aggregates Z, Y. It would be possible to develop a more complete approach by considering the disaggregate data, including the dates of the successive claims and the observed cost by claim. This approach would require the introduction of dynamic duration models with heterogeneity (see Ghysels, Gourieroux, and Jasiak [1998]).
5. See also Dahlby [1985, 1992] and Puelz and Snow [1994] for other studies concerning adverse selection.

References

BEIRLANDT, J., DE MEYER, A.M., DERVEAUX, V., GOOVAERTS, M., LABIE, E., and MAENHUNDT, B. [1991]: "Statistical Risk Evaluation Applied to Belgian Car Insurance," *Insurance: Mathematics and Economics*, 10, 285–302.

BOYER, M., DIONNE, G., and VANASSE, C. [1992]: "Econometric Models of Accident Distribution," in *Contributions to Insurance Economics*, G. Dionne (Ed.), Boston: Kluwer Academic Press, 169–213.

BLOMQUIST, G. [1991]: "Motorist Use of Safety Equipment: Expected Benefits or Risk Incompetence?" *Journal of Risk and Uncertainty*, 4, 135–152.

BUHLMANN, H. [1967]: "Experience Rating and Credibility," *ASTIN Bulletin*, 4, 199–207.

CAMERON, A. and TRIVEDI, P. [1986]: "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, 1, 29–53.

CAMERON, A. and TRIVEDI, P. [1999]: "Essentials of Count Data Regression," forthcoming in *Companion in Econometric Theory*, B. Baltagi (Ed.), New York: Basil Blackwell.

CHANG, J., CHEUNG, C., and KIMSKY, I. [1989]: "On the Derivation of Reinsurance Premiums," *Insurance: Mathematics and Economics*, 8, 137–144.

CHIAPPORI, P.A. and SALANIÉ, B. [1996]: "Empirical Contract Theory: The Case of Insurance Data," CREST DP 9639.

CROCKER, K.J. and SNOW, A. [1986]: "The Efficiency Effects of Categorical Discrimination in the Insurance Industry," *Journal of Political Economy*, 94, 321–344.

DAHLBY, B. [1992]: "Testing for Asymmetrical Information in Canadian Automobile Insurance," in *Contributions to Insurance Economics*, G. Dionne (Ed.), Boston: Kluwer Academic Publishers, 423–444.

DAHLBY, B. [1985]: "Adverse Selection and Statistical Discrimination: An Analysis of Canadian Automobile Insurance," *Journal of Public Economics*, 20, 121–130.

DESJARDINS, D., DIONNE, G., LABERGE-NADEAU, C., and MAAG, U. [1995]: "Medical Conditions, Risk Exposure and Truck Drivers' Accidents: An Analysis with Count Data Regression Models," Thema 9409, forthcoming in *Accident Analysis and Prevention*, 27, 295–305.

DIONNE, G., GAGNÉ, R., GAGNON, F., and VANASSE, C. [1997]: "Debt, Moral Hazard and Airline Safety: An Empirical Evidence," *Journal of Econometrics*, 79, 379–402.

DIONNE, G., GOURIEROUX, C., and VANASSE, C. [1997]: "The Informational Content of Household Decisions with Application to Insurance Under Adverse Selection," CREST DP 9701.

DIONNE, G., GOURIEROUX, C., and VANASSE, C. [1999]: “Evidence of Adverse Selection in Automobile Insurance Markets,” in *Automobile Insurance*, G. Dionne and C. Laberge-Nadeau (Eds.), Boston: Kluwer Academic Publishers, 13–46.

DIONNE, G. and VANASSE, C. [1989]: “A Generalization of Actuarial Automobile Insurance Rating Models: The Negative Binomial Distribution with a Regression Component,” *Astin Bulletin*, 19, 199–212.

DIONNE, G. and VANASSE, C. [1992]: “Automobile Insurance Ratemaking in the Presence of Asymmetrical Information,” *Journal of Applied Econometrics*, 7, 149–165.

DIONNE, G. and VANASSE, C. [1996]: “Une évaluation empirique de la nouvelle tarification de l’assurance automobile au Quebec,” CRT, Montreal, forthcoming in *Econométrie Appliquée, Economica*, C. Gourieroux and C. Montmarquette (Eds.).

GERBER, H. [1990]: *Life Insurance Mathematics*. New York: Springer Verlag.

GERBER, H. and JONES, D. [1973]: “Credibility Formulas with Geometric Weights,” *Proceedings of the Business and Economic Section, American Statistical Association*, 229–230.

GERBER, H. and JONES, D. [1975]: “Credibility Formulas of the Updating Type,” in *Credibility: Theory and Applications*, P. Kahn (Ed.), New York: Academic Press, 89–105.

GHYSELS, E., GOURIEROUX, C., and JASIAK, J. [1997]: “Stochastic Volatility Duration Models,” CREST DP 9746.

GHYSELS, E., HARVEY, A., and RENAULT, E. [1997]: “Stochastic Volatility,” in *Handbook of Econometrics*, vol XIV, G.S. Maddala (Ed.), *Statistical Methods for Finance*, Amsterdam: North-Holland.

GOURIEROUX, C. [1999]: “Statistique de l’assurance,” *Economica*, p. 300.

GOURIEROUX, C. and JASIAK, J. (1998): “Nonlinear Panel Data Models with Dynamic Heterogeneity,” CREST DP 9850.

GOURIEROUX, C. and JASIAK, J. (1999): “Dynamic Factor Models,” CREST DP 9908.

GOURIEROUX, C., MONFORT, A., and TROGNON, A. [1984]: “Pseudo Maximum Likelihood Methods: Application to Poisson Models,” *Econometrica*, 52, 701–721.

GOURIEROUX, C. and SCAILLET, O. [1997]: “Unemployment Insurance and Mortgage,” *Insurance: Mathematics and Economics*, 20, 173–195.

GOURIEROUX, C. and VISSER, M. [1997]: “A Count Data Model with Unobserved Heterogeneity,” *Journal of Econometrics*, 79, 247–268.

HICKMAN, J. and MILLER, R. [1977]: “Notes on Bayesian Graduation,” *Transactions of the Society of Actuaries*, 29, 7–49.

LEMAIRE, J. [1995]: *Bonus-Malus Systems in Automobile Insurance*, Boston: Kluwer Academic Publishers.

NEUHAUS, W. [1988]: “A Bonus-Malus System in Automobile Insurance,” *Insurance: Mathematics and Economics*, 7, 103–112.

PINQUET, J. [1997a]: “Allowance for Cost of Claims in Bonus-Malus Systems,” *ASTIN Bulletin*, 27, 33–57.

PINQUET, J. [1999]: “Allowance for Hidden Information by Heterogenous Models and Applications to Insurance Rating,” in *Automobile Insurance*, G. Dionne and C. Laberge-Nadeau (Eds.), Boston: Kluwer Academic Publishers, 47–78.

PINQUET, J. [1997b]: “Estimating and Testing for Time Dependent Heterogeneity in a Poisson Model,” Thema 9705.

PUELZ, R. and SNOW, A. [1994]: “Evidence on Adverse Selection: Equilibrium Signalling and Cross Subsidization in the Insurance Market,” *Journal of Political Economy*, 42, 236–257.

RICHAUDEAU, D. [1997]: *Modélisation du risque d’accident automobile*, Ph.D. Thesis, Paris I University.

SUNDT, B. [1981a]: “Recursive Credibility Estimation,” *Scandinavian Actuarial Journal*, 3–21.

SUNDT, B. [1981b]: “Credibility Estimators with Geometric Weights,” *Insurance: Mathematics and Economics*, 7, 113–122.

VERRALL, R [1993]: “Graduation by Dynamic Regression Methods,” *JIA*, 120, 153–170.