Deuxième partie

Modèles avec données de panel

Chapitre 3

Fixed Versus Random Effects in Poisson Regression Models for Claim Counts : A Case Study with Motor Insurance

Publié dans *ASTIN Bulletin* 36 : 285-301, 2006
en collaboration avec Michel Denuit

1. Introduction

1.1 Motivation

Risk classification techniques for claim counts have been the topic of many papers appeared in the actuarial literature. Dionne and Vanasse (1989, 1992) used a Poisson model and a Negative Binomial model. Dean et al. (1989) used a Poisson-Inverse Gaussian distribution to fit the number of claims. Recently, Gourieroux and Jasiak (2004) introduced the integer valued autoregressive (INAR) model with unobserved heterogeneity, Denuit and Lang (2004) used Generalised Additive Models, Yip and Yau (2005) presented several parametric Zero-Inflated count distributions and Boucher et al. (2006c) used Zero-Inflated and Hurdle Models.

The mixed Poisson distribution is often used to accounts for unknown characteristics of the driver, influencing the number of accidents reported to the company. When panel data are available, these hidden features can alternatively be captured by an individual heterogeneity term that is constant over time (the standard reference for panel data is Hsiao (2003), while panel data for count variables is treated in Cameron and Trivedi (1998)). This paper aims to confront the two approaches with emphasis on the actual meaning of the estimated parameters in a mixed Poisson regression when random effects and covariates are correlated. In such a case, parameters estimates should be seen has the apparent effects of the covariates on the frequency. Keeping this in mind allows for a better understanding of the resulting price list.

1.2 Agenda

Two kinds of models can be used with longitudinal data : the fixed effects model and the random effects model. Both are briefly described in Section 2. In random effects models, three kinds of heterogeneity distributions are considered in this paper : Gamma, Inverse-Gaussian and Log-Normal. In Section 3, following the work of Mundlak (1978), we link the fixed effects model to the random effects model by a regression on the individual heterogeneity terms. We show that the combination of the fixed effects model with this regression, gives approximately the same results as the random effects model. The final Section 4 concludes.

1.3 Description of the Data

In this paper, we work with a Belgian motor third party liability insurance portfolio comprising 9,894 policies followed for a period of 3 consecutive years (from 1997 to 1999). Thus, we work with 29,682 observations. For each contract, we have informations about the annual number of claims

together with some characteristics of the insured : sex of the driver (man or woman), age of the driver (divided in 3 classes : 17-22, 23-30 and more than 30), power of the vehicle (less than 66kW or more than 66kW) and the size of the city where the insured was living (big, medium or small, based on the number of residents). Figure 1.1 describes the observed annual claim frequency and the distribution of the policyholders according to their characteristics.

2. Panel Data Models

2.1 Presentation

Our portfolio is composed of $N = 9,884$ policyholders. Each policyholder i is observed during $T = 3$ periods. Let $N_{i,t}$ be the number of reported claims for insured i during year t. Such data are called longitudinal data (or panel data). They consist of repeated observations of individual units that are followed over time. Each individual is assumed to be independent of the others but correlation between observations relating to the same individual is permitted. Here, we assume that the number of claims per year obeys to a Poisson distribution with a parameter specific to each policyholder. Specifically, $N_{i,t}$ is assumed to be Poisson distributed with mean $\theta_i\lambda_{i,t}$, $i = 1, ..., N$, $t = 1, ..., T$. The expected annual claim frequency is a product $\theta_i\lambda_{i,t}$ of a static factor θ_i times a dynamic factor $\lambda_{i,t}$. The former accounts for the dependence between observations relating to the same insured. The latter introduces the observable characteristics (that are allowed to vary in time). In general, $\ln\lambda_{i,t}$ is expressed as a linear combination of the observable characteristics, that is $\lambda_{i,t} = \exp(\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}_{i,t})$, where β_0 is the intercept, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ is a vector of regression parameters for explanatory variables $\boldsymbol{x}_{i,t} = (x_{i,t,1}, \ldots, x_{i,t,p})'$.

There are two standard ways of dealing with panel data. In the random effects model, the heterogeneity parameter is treated as a random variable θ_i^{RE} with unit mean. At the portfolio level, the θ_i^{RE}'s are assumed to be independent and identically distributed. In the fixed effects model, the heterogeneity parameter θ_i^{FE} is treated as a parameter to be estimated for each individual. In this case, no intercept enters the model (to ensure identifiability). Conceptually, these two models are quite different. While the fixed effects model makes inferences conditional on the effects present in the sample, the random effects model draws conclusion for the population.

A major difference between the two approaches is that the fixed effects model provides only estimates for the parameters of time varying characteristics, since all the other parameters can been seen as part of the individual term θ_i^{FE}. Further explanations concerning differences between these two models are discussed in Section 3. A standard reference for linear models of panel data is Hsiao

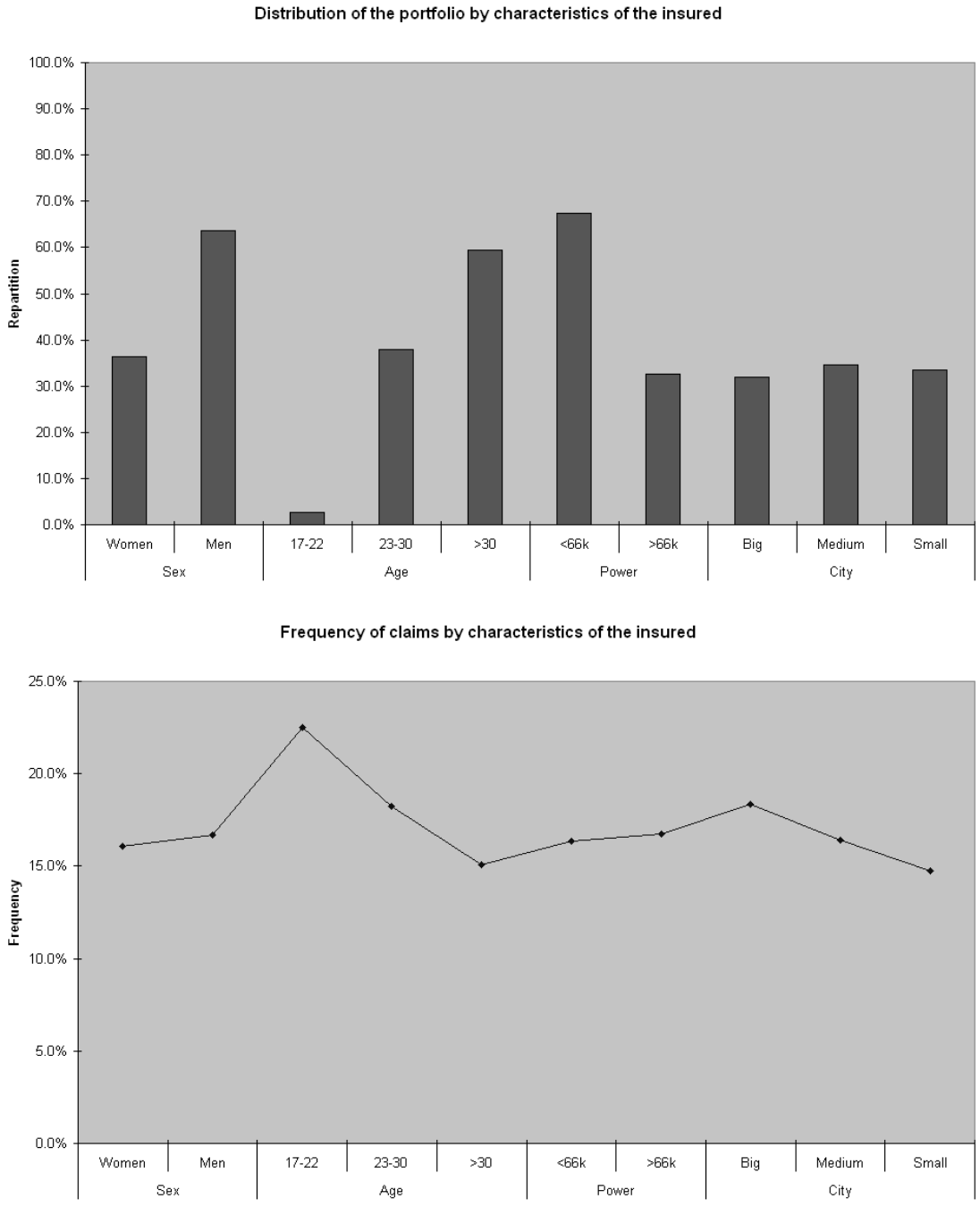


FIG. 1.1 – Description of the motor third party liability portfolio.

(2003), and a good review for count data is provided in Cameron and Trivedi (1998).

2.2 Random Effects Model

2.2.1 Description

In the random effects model (henceforth, quantities relating to the random effects model will be indicated by the superscript RE), θ_i^{RE} is considered as a positive random variable, with probability density function $g(\cdot)$. Given θ_i^{RE}, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. The joint probability function of $N_{i,1}, ..., N_{i,T}$ is thus given by

$$\begin{aligned}
& \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] \\
= & \int_0^\infty \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T} | \boldsymbol{x}_{i,1}, ..., \boldsymbol{x}_{i,T}, \theta_i^{RE}] g(\theta_i^{RE}) d\theta_i^{RE} \\
= & \int_0^\infty \left(\prod_{t=1}^T \Pr[N_{i,t} = n_{i,t} | \boldsymbol{x}_{i,1}, ..., \boldsymbol{x}_{i,T}, \theta_i^{RE}] \right) g(\theta_i^{RE}) d\theta_i^{RE} \\
= & \int_0^\infty \left(\prod_{t=1}^T \exp(-\theta_i^{RE} \lambda_{i,t}^{RE}) \frac{(\theta_i^{RE} \lambda_{i,t}^{RE})^{n_{i,t}}}{n_{i,t}!} \right) g(\theta_i^{RE}) d\theta_i^{RE}. \qquad (2.1)
\end{aligned}$$

Estimation of parameters is performed using maximum likelihood estimators or moment techniques. The GEE method of Liang and Zeger (1986) can be a solution to account for the dependence between each observation of the same insured as shown in Denuit et al. (2003). However, for this paper, we will restrict ourselves to maximum likelihood estimates for all models analysed.

2.2.2 Poisson-Gamma Model

If θ_i^{RE} follows the Gamma distribution with mean 1 and variance $\frac{1}{\nu}$, the joint probability function of $N_{i,1}, ..., N_{i,T}$ writes

$$\begin{aligned}
& \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] \\
& = \left(\prod_{t=1}^T \frac{(\lambda_{i,t}^{RE})^{n_{i,t}}}{n_{i,t}!} \right) \frac{\Gamma(n_{i,\bullet} + \nu)}{\Gamma(\nu)} \left(\frac{\nu}{\sum_{i=1}^T \lambda_{i,t}^{RE} + \nu} \right)^\nu \left(\sum_{i=1}^T \lambda_{i,t}^{RE} + \nu \right)^{-n_{i,\bullet}}, \qquad (2.2)
\end{aligned}$$

where $n_{i,\bullet} = \sum_{t=1}^T n_{i,t}$. In this case,

$$\mathbb{E}[N_{i,t}] = \lambda_{i,t}^{RE} < \mathbb{V}[N_{i,t}] = \lambda_{i,t}^{RE} + (\lambda_{i,t}^{RE})^2 / \nu.$$

Maximum likelihood estimations of parameters and variances can be obtained as follows. The first order conditions for parameters β_0^{RE}, $\boldsymbol{\beta}^{RE}$ and ν lead to the system

$$\sum_{i=1}^{n}\sum_{t=1}^{T} n_{i,t} - \lambda_{i,t}\frac{\sum_t n_{i,t}+\nu}{\sum_t \lambda_{i,t}+\nu} = 0 \tag{2.3}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \boldsymbol{x}_{i,t}\left(n_{i,t} - \lambda_{i,t}\frac{\sum_t n_{i,t}+\nu}{\sum_t \lambda_{i,t}+\nu}\right) = \mathbf{0} \tag{2.4}$$

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n_{i,\bullet}-1}\frac{1}{j+\nu}\right) - \log\left(1+\frac{\sum_t \lambda_{i,t}}{\nu}\right) + \sum_t \frac{\lambda_{i,t}+n_{i,t}}{\sum_t \lambda_{i,t}+\nu} = 0. \tag{2.5}$$

Numerical procedures can then be used to solve these equations.

2.2.3 Poisson-Inverse Gaussian Model

The Inverse Gaussian distribution is another good candidate to model the heterogeneity parameter (Willmot (1987), Dean et al. (1989) and Tremblay (1992) for an application with insurance data). Shoukri et al. (2004) showed that the Poisson with Inverse-Gaussian heterogeneity of mean and variance equal to 1 and τ respectively, has the joint probability function

$$\begin{aligned}&\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}]\\ &= \left(\prod_{t=1}^{T}\frac{(\lambda_{i,t}^{RE})^{n_{i,t}}}{n_{i,t}!}\right)\left(\frac{2}{\pi\tau}\right)^{0.5} e^{1/\tau}\left(1+2\tau\sum_{t=1}^{T}\lambda_{i,t}^{RE}\right)^{-s_i/2} K_{s_i}(z_i)\end{aligned} \tag{2.6}$$

where $K_j(.)$ is the modified Bessel function of the second kind, $s_i = n_{i,\bullet} - 0.5$ and

$$z_i = \frac{1}{\tau}\sqrt{1+2\tau\sum_{t=1}^{T}\lambda_{i,t}^{RE}}.$$

The modified Bessel function of the second kind has some useful properties that can be used to find the maximum likelihood estimators (for more details, see Shoukri et al. (2004)).

Now, $\mathbb{V}[N_{i,t}] = \lambda_{i,t}^{RE} + \tau(\lambda_{i,t}^{RE})^2$. The maximum likelihood estimators of β_0^{RE} and $\boldsymbol{\beta}^{RE}$ solve

$$\sum_{i=1}^{n}\sum_{t=1}^{T} n_{i,t} - \lambda_{i,t}M\left(n_{i,\bullet}\right) = 0 \tag{2.7}$$

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \boldsymbol{x}_{i,t}\left(n_{i,t} - \lambda_{i,t}M\left(n_{i,\bullet}\right)\right) = \mathbf{0}, \tag{2.8}$$

where the function $M(.)$ can be expressed as the ratio of the modified Bessel function of the second kind

$$\frac{K_{n_{i,\bullet}+1/2}(a)}{K_{n_{i,\bullet}-1/2}(a)} = M(n_{i,\bullet})\sqrt{1+2\tau\sum_{t=1}^{T}\lambda_{i,t}^{RE}}.$$

The estimation of the parameter τ can be found by solving

$$-\frac{1}{2\tau}-\frac{1}{\tau^2}-\frac{s_i\sum_t\lambda_{i,t}}{1+2\tau\sum\lambda_{i,t}}+\frac{\partial\log K_{s_i}(z_i)}{\partial\tau}=0 \tag{2.9}$$

where the derivative of the function K is equal to :

$$\frac{\partial\log K_{s_i}(z_i)}{\partial z_i} = -M(n_{i,\bullet})\sqrt{1+2\tau\sum_t\lambda_{i,t}}+\frac{(n_{i,\bullet}-\frac{1}{2})\tau}{\sqrt{1+2\tau\sum_t\lambda_{i,t}}} \tag{2.10}$$

$$\frac{\partial z_i}{\partial\tau} = \frac{\sum_t\lambda_{i,t}}{\tau\sqrt{1+2\tau\sum_t\lambda_{i,t}}}-\frac{\sqrt{1+2\tau\sum_t\lambda_{i,t}}}{\tau^2}. \tag{2.11}$$

Again, numerical procedures are needed to obtain the solutions.

2.2.4 Poisson-Log Normal Model

In biostatistical circles, the Poisson Log-Normal model is often used, after Hinde (1982). In this case, the error term can be expressed as $\theta_i^{RE}=\exp(\epsilon_i)$ for some Gaussian noise ϵ_i. From this, the mean parameter has the form $\exp(\boldsymbol{x}_{i,t}'\boldsymbol{\beta}^{RE}+\epsilon_i)=\gamma_{i,t}$, with ϵ_i following a Gaussian distribution with mean $\mu=-\sigma^2/2$ (a simplifying normalization) and variance σ^2. In this case,

$$\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}]=\int_{-\infty}^{\infty}\left(\prod_{t=1}^{T}\frac{e^{-\gamma_{i,t}}(\gamma_{i,t})^{n_{i,t}}}{n_{i,t}!}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right)d\epsilon_i. \tag{2.12}$$

Numerical procedures can be used to find maximum likelihood estimates. Routines are now available in standard statistical packages, such as SAS (with the NLMIXED procedure). Now, $\mathbb{E}[N_{i,t}]=\lambda_{i,t}^{RE}$ and $\mathbb{V}[N_{i,t}]=\lambda_{i,t}^{RE}+(e^{\sigma^2}-1)(\lambda_{i,t}^{RE})^2$.

The Poisson-Log Normal model is interesting because it has a natural interpretation (see, e.g., Winkelmann (2003a)). The error term ϵ_i is often considered as a factor that captures the effects of hidden exogeneous variables. If there are many hidden variables, and if these variables are independent, then central limit theorems can be invoked in order to establish the normality of ϵ_i.

2.2.5 Numerical Example

Estimations of the parameters for the 3 random effects models are displayed in Table 2.1. Sex of the driver and power of the car have been removed from all models since they were not statistically

Variable	Parameter	Gamma	Inv. Gaussian	Log Normal
Intercept	-	-2.0059 (0.034)	-2.0063 (0.034)	-2.0064 (0.034)
Age	$17-22$	0.4009 (0.088)	0.4030 (0.088)	0.4034 (0.088)
	$23-30$	0.1964 (0.034)	0.1983 (0.034)	0.1988 (0.034)
	>30	0	0	0
City	Big	0.2337 (0.041)	0.2345 (0.041)	0.2346 (0.041)
	Medium	0.1106 (0.041)	0.1100 (0.041)	0.1098 (0.041)
	Small	0	0	0
ν	-	2.6638 (0.314)	-	-
τ	-	-	0.3940 (0.048)	-
σ^2	-	-	-	0.3363 (0.036)
Log-Likelihood	-	-12,937.60	-12,936.30	-12,936.02

TAB. 2.1 – Parameter Estimates $\widehat{\boldsymbol{\beta}}^{RE}$ of the Poisson Random Effects Models.

significant, with respective p-values of approximately 0.38 and 0.12 for all models (specifically, p-values of 0.3841 and 0.1163 in the Poisson-Gamma model, 0.3861 and 0.1205 in the Poisson-Inverse Gaussian model, 0.3843 and 0.1205 in the Poisson-LogNormal model). However, other variables have considerable impact such as the age of the driver or the size of the city where the insured lives. From Table 2.1, we see that young drivers and policyholders living in big cities exhibit higher expected claim frequencies. All models seem to have approximately the same quality of fit since their log-likelihood are almost equal for the same number of parameters.

2.3 Fixed Effects Model

2.3.1 Description

In the fixed effects model, all characteristics that are not time-varying are captured by the individual heterogeneity term θ_i^{FE}. In our case, the intercept β_0 has to be removed (and is included in θ_i^{FE}). As sex of the driver has been removed from the model, the remaining explanatory variables do vary with time and enter the fixed effects model.

The Poisson fixed effects model has been proposed by Palmgren (1981) and Hausman et al. (1984). The standard way of evaluating the parameters of this model is the conditional maximum likelihood of Andersen (1970). The idea of the conditional method is to obtain an estimator of $\boldsymbol{\beta}^{FE}$

without having to estimate each θ_i^{FE}.

As proved in Cameron and Trivedi (1998), the maximum likelihood and conditional maximum likelihood estimation methods always yield identical estimates for covariates parameter $\boldsymbol{\beta}^{FE}$ in case of Poisson distribution. Specifically, the estimated parameters solve

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\boldsymbol{x}_{i,t}\left(n_{i,t}-\lambda_{i,t}^{FE}\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}^{FE}}\right)=\mathbf{0}. \tag{2.13}$$

Note that only insureds with varying caracteristics (and at least one claim) are used in the estimation of $\boldsymbol{\beta}^{FE}$. Estimates of θ_i^{FE} can then be obtained from

$$\theta_i^{FE}=\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}^{FE}}. \tag{2.14}$$

Both estimates of $\boldsymbol{\beta}^{FE}$ and θ_i^{FE} are consistent when $T\to\infty$ and $N\to\infty$, but only the estimate of $\boldsymbol{\beta}^{FE}$ is consistent for fixed T and $N\to\infty$, as for insurance data.

2.3.2 Numerical Example

Application of the fixed effects Poisson model to the Belgian motor portfolio leads to the parameters estimates of $\boldsymbol{\beta}^{FE}$ displayed in Table 2.2. Some interesting conclusions can be drawn from the fixed effects model. The results are quite different from those obtained with the random effects models. Indeed, young drivers can be seen as better drivers than older ones and insureds coming from big cities now seem to be better drivers. It is worth to stress that the estimated parameters in Table 2.2 have to be thought in the sense of fixed effects model, where all individual impacts are removed.

The true effect of young age and living in large cities is thus to decrease the annual expected claim frequency. The apparent higher risk that is often found to be associated with these characteristics in empirical studies then results from their association with dangerous individual characteristics.

The high values of the standard errors suggest that almost all parameter are not statistically different from zero. This is nevertheless not a problem here, since our aim is to show that a fixed effects model followed by a regression of the θ_i^{FE}'s on the observable characteristics produces almost the same results than a random effects model.

2.4 Fixed or Random Effects Model ?

Comparaison between equations (2.13) and (2.4) or (2.8) leads to the conclusion that there are no differences between fixed or random effects when T is large enough. However, for fixed small

Variable	Parameter	Estimate (Std. err.)
Age	$17-22$	$-0.4754(0.205)$
	$23-30$	$-0.1603(0.126)$
	>30	0
City	Big	$-0.7945(0.523)$
	Medium	$-0.3418(0.530)$
	Small	0
Log-Likelihood		-7803.577

TAB. 2.2 – Parameter Estimates $\widehat{\boldsymbol{\beta}}^{FE}$ of the Poisson Fixed Effects Model.

T, parameters estimates for these two models can be significantly different (as it can be seen from Tables 2.1-2.2).

A reason for such a difference between parameters estimates comes from the construction of the random effects model. Indeed, in the development of equation (2.1), we have made the following crucial assumption : the random effects θ_i^{RE} are independent and identically distributed. This means that the conditional probability density function of θ_i^{RE} given $\boldsymbol{x}_{i,t}$ equals $g(\cdot)$, that does not depend on $\boldsymbol{x}_{i,t}$. If the distribution of θ_i^{RE} depends on the $\boldsymbol{x}_{i,t}$'s, the parameters estimates $\widehat{\boldsymbol{\beta}}^{RE}$ may be inconsistent and should not be used, as shown by Mundlak (1978).

Figure 2.1 shows us the distribution of the θ_i^{FE} (mean and 0 to 95th percentile) according to the characteristics of the insured. Clearly, we can see that the distribution of the heterogeneity term varies with the characteristics of the insured. The dispersion is much more important in the least dangerous classes, like Age 17-22 and Big cities. On average, the θ_i^{FE}'s are larger there, which explains the apparent riskiness in Table 2.1. The heterogeneity is not identically distributed, which can cause an inconsistency in the evaluation of the parameters β^{RE} of the random effects model.

Application of Hausman test to our data leads, without surprise, to the rejection of the null hypothesis of uncorrelation between regressors and random effects (p-value of less than 0.01% for a chi-square distribution with 4 degrees of freedom). This result means that the heterogeneity term is not identically distributed accross insureds.

The fixed effects model is preferred in cases where conclusions have to be made on the sample, while the interests of random effects model are on the overall population. For insurance data, fixed effects model cannot be used since annual premiums cannot be calculated for new policyholders.

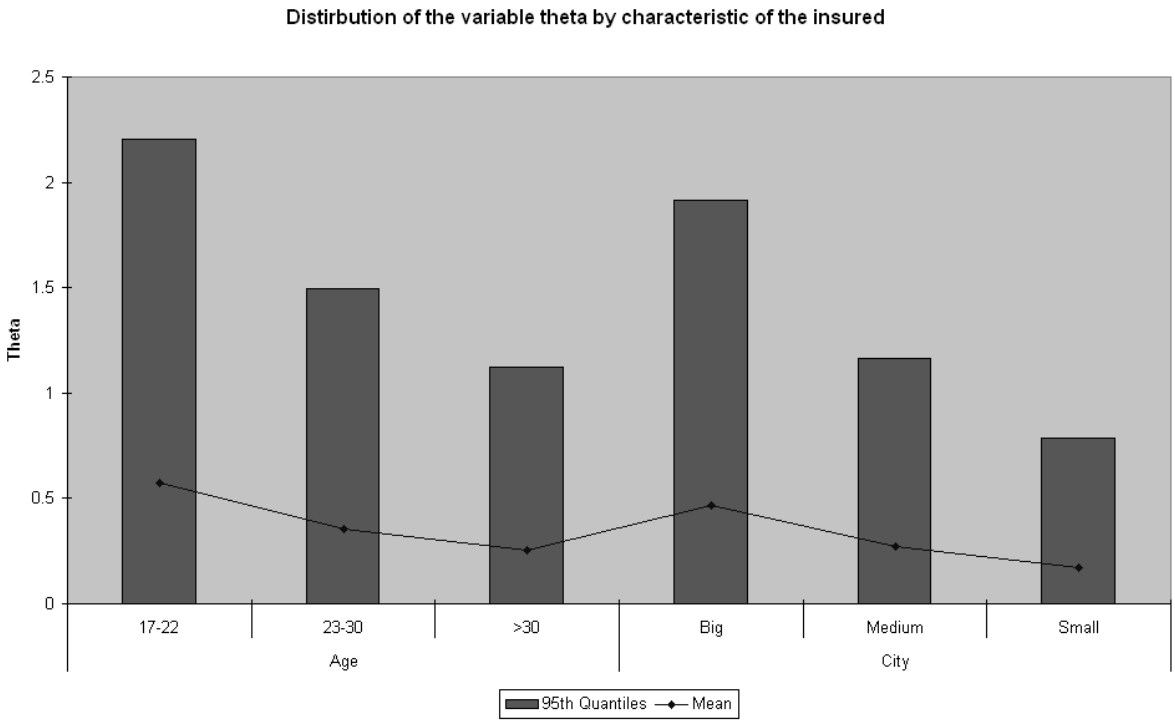


FIG. 2.1 – Distribution of the θ_i^{FE}'s according to the observable characteristics.

Expected claim frequencies can be computed only for policyholders in the portfolio for several years. Moreover, estimates for the θ_i^{FE}'s based on just a few observations must be considered with caution. Additionnaly, fixed effects models are difficultly handled by insurance companies since too many individual effects had to be considered. Random effects models should then be preferred, but, as we mentioned, some theorical aspects prohibit its use since estimates of the parameters are biased when heterogeneity is not independent from the regressors. The purpose of the next section is to legitimate the use of the random effects models for insurance ratemaking, provided parameters are estimated and interpreted with care.

3. Regression of the θ_i^{FE}'s on the observable characteristics

3.1 Residual Heterogeneity

As noted by Pinquet (2000), the random effects are often correlated with covariates for insurance data. Therefore, they relate to some residual heterogeneity, that is, orthogonal to the observable variables. The aim of this section is to demonstrate on a basis of the Belgian data set that a fixed effect Poisson regression followed with a regression of the resulting θ_i^{FE}'s on the observable characteristics gives almost the same values for the regression coefficients than a random effect Poisson regression. This legitimates the use of random effects techniques in actuarial science. We also consider mixed effects model with non identically distributed θ_i^{RE}, and we reach a similar conclusion.

Specifically, let us now explain the $\hat{\theta}_i^{FE}$'s as a function of covariates. To this end, we consider the $\hat{\theta}_i^{FE}$'s as realizations from some probability density function $h(\mu_i, \tau)$ with mean $\mu_i = \exp(\hat{\boldsymbol{x}}_i'\boldsymbol{\delta})$ and variance τ. The term $\hat{\boldsymbol{x}}_i$ is an adaptation of covariates $\boldsymbol{x}_{i,t}$ for $t = 1, ..., T$, since there is only one indivual heterogeneity term for all periods t (here, we took $\hat{\boldsymbol{x}}_i = x_{i,2}$). We will take for h the Gamma, Inverse Gaussian and LogNormal densities (also considered in the random effects model).

We saw on Figure 2.1 that the distribution of the θ_i^{FE}'s was influenced by the observable characteristics. Another way of explaining the random effects model (second approximation) is to consider non-equally distributed heterogeneity. In consequence, instead of assuming that the θ_i^{RE}'s are independent and identically distributed with mean 1 and variance τ, we use an other model. The evaluation is done with known λ_i^{FE} and we assume that the θ_i^{FE}'s are distributed with a mean of $\mu_i = \exp(\hat{\boldsymbol{x}}_i'\boldsymbol{\delta})$ and variance τ_i. Formally, we now consider that

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \int_0^\infty \left(\prod_{t=1}^{T} \frac{(\lambda_{i,t}^{FE}\theta_i)^{n_{i,t}} \exp(-\lambda_{i,t}^{FE}\theta_i)}{n_{i,t}!} \right) f(\theta_i|\mu_i, \tau_i) d\theta_i, \quad (3.1)$$

where $f(\cdot|\mu_i, \tau_i)$ denotes some probability density function with mean $\mu_i = \exp(\hat{\boldsymbol{x}}_i'\boldsymbol{\delta})$ and variance τ_i. Again, we consider for f the Gamma, Inverse Gaussian and LogNormal densities.

If the two-step procedure based on the θ_i^{FE}'s, or the one-step procedure based on heterogeneous θ_i^{RE}, is coherent with the random effect model, we expect that

$$\widehat{\boldsymbol{\beta}}^{RE} \approx \widehat{\boldsymbol{\beta}}^{FE} + \widehat{\boldsymbol{\delta}}. \tag{3.2}$$

As proved by Mundlak (1978), the fact that there exists a correlation between θ_i^{RE} and the covariates causes a bias on the estimation of the β^{RE}. However, despite the presence of this bias, it is still possible to use the parameter estimates for the premium calculation if the heterogeneity term is only considered as residual heterogeneity. In consequence, these estimates represent the apparent effects on the frequency of claims and not the real effect, since we must use $(\widehat{\boldsymbol{\beta}}^{FE} + \widehat{\boldsymbol{\delta}})$ instead of $\widehat{\boldsymbol{\beta}}^{FE}$. For insurance ratemaking, this distinction does not really matter since apparent effect is the interest when some important classification variables are missing.

Remark 3.1. The equation used to construct a regression analysis on the individual fixed effects is based on equation (2.14). The individual fixed effects are expressed as the ratio of the sum of the number of claims in the T years on the sum of the $\lambda_{i,t}^{FE}$ for the same period. Consequently, there is a significant presence of zero value for the θ_i^{FE}, which causes a problem for a regression analysis. In consequence, weighted regression seems appropriate for modelling the individual fixed effects. However, weighted observations without claim need to be removed from the dataset since it is not possible to work with a zero-valued observation in a regression analysis of heterogeneity. Still because of the weighted regression, it becomes impossible to have the same dispersion parameter for random effects models and its first approximation, since some random variations are removed by using average values.

3.2 Gamma Heterogeneity

Let us begin with Gamma distributed θ_i^{FE}'s. Since the Gamma distribution is a member of the exponential family, the estimation of $\boldsymbol{\delta}$ is based on GLM (see McCullagh and Nelder (1989)) and solves :

$$\begin{aligned}\sum_{i=1}^{n}\frac{1}{\omega_i}(\theta_i^{FE}-\mu_i)\frac{\partial\mu_i}{\partial\boldsymbol{\delta}} &= \sum_{i=1}^{n}\frac{\nu}{\mu_i}(\theta_i^{FE}-\mu_i)\hat{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{n}\frac{\nu}{\mu_i}\left(\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}^{FE}}-\mu_i\right)\hat{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{n}\frac{\nu}{\sum_t \lambda_{i,t}^{FE}\mu_i}\left(\sum_t n_{i,t}-\sum_t \lambda_{i,t}^{FE}\mu_i\right)\hat{\boldsymbol{x}}_i=\mathbf{0} \end{aligned} \tag{3.3}$$

where $\omega_i=\mu_i^2/\nu$ is the variance of the distribution and μ_i is the mean function that is expressed as $\exp(\hat{x}_i'\delta)$. Since $N_{i,t}$ is a discrete variable, the resulting first order condition has the form of a Gamma distribution for $\sum_t n_{i,t}$, with mean $\sum_t \lambda_{i,t}^{FE}\mu_i$ and variance proportional to the square of the mean.

If we allow for non identically distributed random effect, we get the following contribution for policyholder i to the likelihood in the random effects model :

$$\begin{aligned}&\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}] \\ =\ &\int_0^\infty\left(\prod_{t=1}^{T}\frac{(\lambda_{i,t}^{FE}\theta_i)^{n_{i,t}}\exp(-\lambda_{i,t}^{FE}\theta_i)}{n_{i,t}!}\right)\left(\frac{\theta_i\nu}{\mu_i}\right)^{\nu}(\theta_i)^{-1}\frac{\exp(\frac{-\theta_i\nu}{\mu_i})}{\Gamma(\nu)}d\theta_i \\ =\ &\left(\prod_{t=1}^{T}\frac{(\lambda_{i,t}^{FE}\mu_i)^{n_{i,t}}}{n_{i,t}!}\right)\frac{\nu^\nu}{\Gamma(\nu)}\frac{\Gamma(n_{i,\bullet}+\nu)}{(\mu_i\sum_{t=1}^{T}\lambda_{i,t}^{FE}+\nu)^{n_{i,\bullet}+\nu}}. \end{aligned} \tag{3.4}$$

Only two observations have been removed from the weighted regression. The parameters of the first model have been found by a two-step procedure. Firstly, the individual heterogeneity parameters θ_i^{FE} have been evaluated using equation (2.14). Afterwards, a weighted Gamma regression has been performed to find estimates of $\boldsymbol{\delta}$. For the second model, equation (3.4) has been applied with $\lambda_{i,t}^{FE}$ known from application of the fixed effects model.

We can see in Table 3.1 that the parameter estimates of $\boldsymbol{\delta}$ for the first and second models are approximately equal to the difference between the random and the fixed effects model.

3.3 Inverse Gaussian Heterogeneity

Once again, instead of supposing that the heterogeneity term is independant of the covariates, we suppose that the heterogeneity term has Inverse Gaussian distribution with mean μ_i and a corresponding variance ω_i that is equal to $\mu_i^3\tau$. Inverse Gaussian distribution can also be estimated using first-order condition of the GLM (see McCullagh and Nelder (1989)). The estimator of $\boldsymbol{\delta}$ then

Variable	Parameter	RE-FE Difference	1st model	2nd model
Intercept	-	-2.0128 (0.034)	-1.9971 (0.044)	-2.0062 (0.034)
Age	$17-22$	0.8763 (0.038)	0.9016 (0.132)	0.8625 (0.085)
	$23-30$	0.3567 (0.018)	0.3339 (0.047)	0.3533 (0.034)
	>30	0	0	0
City	Big	1.0282 (0.234)	1.0223 (0.056)	1.0369 (0.041)
	Medium	0.4524 (0.166)	0.425 (0.055)	0.4570 (0.041)
	Small	0	0	0
Dispersion	ν	2.6638 (0.314)	0.1975 (0.065)	2.6437 (0.310)

TAB. 3.1 – Parameter Estimates for Gamma Heterogeneity

solves

$$\begin{aligned}\sum_{i=1}^{n}\frac{1}{\omega_i}(\theta_i^{FE}-\mu_i)\frac{\partial\mu_i}{\partial\boldsymbol{\delta}} &= \sum_{i=1}^{n}\frac{\tau}{\mu_i^2}(\theta_i^{FE}-\mu_i)\hat{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{n}\frac{\tau}{\mu_i^2}\left(\frac{\sum_t n_{i,t}}{\sum_t \lambda_{i,t}^{FE}}-\mu_i\right)\hat{\boldsymbol{x}}_i \\ &= \sum_{i=1}^{n}\frac{\tau}{\sum_t \lambda_{i,t}^{FE}\mu_i^2}\left(\sum_t n_{i,t}-\sum_t \lambda_{i,t}^{FE}\mu_i\right)\hat{\boldsymbol{x}}_i=\mathbf{0}. \end{aligned} \tag{3.5}$$

If we allows for random effects with different means, the contribution of policyholder i to the likelihood is

$$\begin{aligned}&\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}] \\ =\ &\int_0^\infty\left(\prod_{t=1}^{T}\frac{(\lambda_{i,t}^{FE}\theta_i)^{n_{i,t}}\exp(-\lambda_{i,t}^{FE}\theta_i)}{n_{i,t}!}\right)(2\pi\tau\theta_i^3)^{-1/2}\exp\left(\frac{-(\theta_i-\mu_i)^2}{2\mu_i^2\theta_i\tau}\right)d\theta_i \\ =\ &\left(\prod_{t=1}^{T}\frac{(\lambda_{i,t}^{FE}\mu_i)^{n_{i,t}}}{n_{i,t}!}\right)\left(\frac{2}{\pi\tau\mu_i}\right)^{0.5}\exp(\frac{1}{\tau\mu_i}) \\ &\times\left(1+2(\tau\mu_i)\sum_{t=1}^{T}\lambda_{i,t}\mu_i\right)^{-s_i/2}K_{s_i}(z_i)\end{aligned} \tag{3.6}$$

where $K_j(.)$ is the modified Bessel function of the second kind, $s_i=n_{i,\bullet}-0.5$ and

$$z_i=\frac{1}{\tau}\sqrt{1+2(\tau\mu_i)\sum_{t=1}^{T}\lambda_{i,t}^{FE}\mu_i}.$$

Variable	Parameter	RE-FE Difference	1st model	2nd model
Intercept	-	-2.0127 (0.034)	-1.9888 (0.033)	-2.0098 (0.033)
Age	$17-22$	0.8784 (0.037)	0.8758 (0.169)	0.8431 (0.090)
	$23-30$	0.3585 (0.018)	0.3101 (0.045)	0.3376 (0.034)
	>30	0	0	0
City	Big	1.0289 (0.234)	1.0097 (0.059)	1.0271 (0.041)
	Medium	0.4517 (0.166)	0.4456 (0.047)	0.4603 (0.040)
	Small	0	0	0
Dispersion	τ	0.3940 (0.048)	4.2234 (0.704)	1.090 (0.158)

TAB. 3.2 – Parameter Estimates for Inverse Gaussian Heterogeneity.

Table 3.2 shows the results of these models on our insurance data. As expected, the results are quite the same as those obtained with the Gamma heterogeneity models. Even the intercept of the first approximation model has approximately the same value, due to the presence of high values in the data.

One difference between models is the parameter τ. This difference comes from the fact that $\tau_i^{RE_1} = \tau^{RE_2}\mu_i$, where RE_1 and RE_2 are the first and second approximation models. Since the value of μ_i is 0.3326, we can link the two estimated parameters τ by the following approximation : $0.3940 \approx 1.090 \times 0.3326 = 0.3625$.

3.4 Log-Normal Heterogeneity

Let us now assume that $\theta_i^{FE} = \exp(\epsilon_i)$, where ϵ_i is following a Gaussian distribution with mean $\mu_i = \boldsymbol{x}_i'\boldsymbol{\delta} - \frac{\sigma^2}{2}$ and variance $\omega_i = \sigma^2$. Obviously, the Gaussian distribution can be solved using first-order condition of the GLM models (see McCullagh and Nelder (1989)). In consequence, the estimation of $\boldsymbol{\delta}$ solves

$$\begin{aligned}
\sum_{i=1}^{n} \frac{1}{\omega_i}\big(\log(\theta_i^{FE}) - \log(\mu_i)\big)\frac{\partial \mu_i}{\partial \boldsymbol{\delta}} &= \sum_{i=1}^{n} \frac{1}{\sigma^2}\left(\log(\theta_i^{FE}) - \log(\mu_i)\right)\hat{\boldsymbol{x}}_i \\
&= \sum_{i=1}^{n} \frac{1}{\sigma^2}\left(\log(\sum_t n_{i,t}) - \log(\sum_t \lambda_{i,t}^{FE}) - \log(\mu_i)\right)\hat{\boldsymbol{x}}_i \\
&= \sum_{i=1}^{n} \frac{1}{\sigma^2}\left(\log(\sum_t n_{i,t}) - \log(\sum_t \lambda_{i,t}^{FE}\mu_i)\right)\hat{\boldsymbol{x}}_i = \mathbf{0}. \qquad (3.7)
\end{aligned}$$

Variable	Parameter	RE-FE Difference	1st model	2nd model
Intercept	-	-2.0129 (0.034)	-2.0037 (0.044)	-2.0210 (0.034)
Age	$17-22$	0.8788 (0.037)	0.8793 (0.132)	0.8668 (0.086)
	$23-30$	0.3591 (0.018)	0.3295 (0.047)	0.3560 (0.034)
	>30	0	0	0
City	Big	1.0292 (0.234)	1.0209 (0.056)	1.0381 (0.041)
	Medium	0.4515 (0.166)	0.4539 (0.055)	0.4562 (0.041)
	Small	0	0	0
Dispersion	σ^2	0.3363 (0.036)	2.2640 (0.3773)	0.3422 (0.036)

TAB. 3.3 – Parameter Estimates for Log-Normal Heterogeneity.

Allowing for non identically distributed random effects, the joint probability density function of the claim numbers for insured i is

$$\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}]=\int_{-\infty}^{\infty}\left(\prod_{t=1}^{T}\frac{e^{-\gamma_i}\gamma_i^{n_i}}{n_i!}\right)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(\epsilon_i-\mu_i)^2}{2\sigma^2}\right)d\epsilon_i \tag{3.8}$$

with $\gamma_i=\exp(\boldsymbol{x}_i'\boldsymbol{\beta}+\epsilon_i)$. Closed form of the Poisson-Log-Normal distribution is not possible. However, by numerical approximations or by the NLMIXED procedure in SAS, it is possible to find estimators of $\boldsymbol{\delta}$.

The numerical results are displayed in Table 3.3. Approximation of a weighted regression gives interesting results that are similar to those obtain with the Gamma and the Inverse-Gaussian distributions.

4. Concluding Remarks

The results brought here give us the legitimity to use random effects models even if there exists a correlation between the regressors and the heterogeneity. The parameter estimates do not identify the impact of these regressors on the premium but only the apparent effects. Since it is usually the interest of the actuary in ratemaking, there is no problem with this interpretation. However, such a correlation indicates clearly that a correction should be done to obtain a more accurate model. Especially, the apparent high risk of young drivers should deserve some attention. The analysis conducted in this paper shows that the fixed effects are very heterogeneous for these individuals. Instead of penalizing these insureds in the a priori ratemaking, an appropriate bonus-malus scheme

could be designed. Merit rating systems improve the fairness of the tariff in that respect.

We have focused our attention on the Poisson distribution since it is commonly used in practice for risk classification. However, other count distributions, such as the Negative Binomial distribution (see Hausman et al. (1984) or Allison and Waterman (2002) for an alternative fixed effects model) or the Zero-Inflated models also deserve consideration for the analysis of claim frequencies. A similar analysis could be performed for these regression models, too. Moreover, continuous probability models, such as the Gamma or the Log-Normal distributions are routinely used to model the amount of claims. Again, a study in the vein of the present one could be performed for claim severities.

Chapitre 4

Models of Insurance Claim Counts with Time Dependence based on Generalisations of Poisson and Negative Binomial Distributions

Publication prochaine dans *Variance*
en collaboration avec Michel Denuit et Montserrat Guillén

1. Introduction

Modeling the number of claims is an essential part of insurance pricing. Count regression analysis allows idenfication of risk factors and prediction of the expected frequency given the characteristics of the policyholders. A usual way to calculate the premium is to obtain the conditional expectation of the number of claims given the risk characteristics and to combine it with the expected claim amount. Some insurers may also consider experience rating when setting the premiums, so that the number of claims reported in the past, can be used to improve the estimation of the conditional expectation of the number of claims for the following year.

The literature on count regression analysis has grown considerably in the past years. Simultaneously, insurers accumulate longitudinal information on their policyholders. In this paper we want to address panel count data models in the context of insurance, so that we can see the advantages of using the information on each policyholder along time for modeling the number of claims. The existing literature has mainly advocated the use of the classical Poisson regression approach, but little has been said on other model alternatives and on model selection. Although in the insurance context, the main interest is to obtain the conditional expectation of the number of claims, we argue that new panel data models presented here that allow for time dependence between observations are closer to the data generating process that one can find in practice. Moreover, closed form expressions for future premiums given the past observations and model selection will definitely help practitioners to find suitable alternatives for modeling insurance portfolios that have accumulated some years of history.

These panel data models generalize the Poisson and the Negative Binomial distributions into 6 different distributions, based on Bayesian and frequentist modeling. Some of those models are first introduced to actuarial sciences. Each section presents a different model for panel data, where the interpretation of the model, the probability distribution, the properties of the model and its first moments are shown. In some sections, it is shown why some intuitive models (past experience as regressors, multivariate distributions or copula models) involving time dependence cannot be used to model the number of reported claims. Statistical tests to compare the nested models are explained and a Vuong test is used to compares the fitting of non-nested models. Our results show that the random effects models have a better fit than the alternative models presented here and have more flexibility in computing next years premium, that it is not only based on autocorrelation but on Bayesian modeling.

Variable	Description
v1	equals 1 for women and 0 for men
v2	equals 1 if the client has been in the company between 3 and 5 years
v3	equals 1 if the client has been in the company for more than 5 years
v4	equals 1 if the insured is 30 years old or younger
v5	equals 1 if power is larger or equal to 5500 cc

TAB. 1.1 – Exogenous variables

1.1 Agenda

Section 2 presents the Poisson and the Negative Binomial distributions in time independence situation. In section 3, a Bayesian model for time dependence, where the Poisson and Negative Binomial distributions are modeled by a common random effects, is shown. Models where past experience is modeled as covariate are explained in section 4. Section 5 reviews the INAR(1) models while section 6 presents common shock models. Reported claims modeled by copulas are introduced in section 7. Section 8 compares a priori premiums and predictive distribution of all models analysed while section 9 presents some specification tests that can be used to compare models.

1.2 Data used and Estimation

In this paper, we worked with a sample of the automobile portfolio of a major company operating in Spain. Only private use cars have been considered in this sample. The panel data contains information from 1991 to 1998. Our sample contains 15,179 policyholders that stay in the company for seven complete periods, resulting in 106,253 insurance contracts.

We have 5 exogeneous variables, described in the table 1.1, that are kept in the panel plus the yearly number of accidents. For every policy we have the initial information at the beginning of the period and the total number of claims at fault that took place whithin this yearly period. The average claim frequency is 6.8412%.

In this paper, these exogenous variables are used to model the parameters of the distributions. The characteristics of the insureds are expressed throught functions such as $h(\beta_0 + x'_{i,t}\beta)$, where β_0 is the intercept, $\beta' = (\beta_1, ..., \beta_p)$ is a vector of regression parameters for explanatory variables

$x_{i,t} = (x_{i,t,1}, ..., x_{i,t,p})$. Subscript t may be removed in case where the covariates do not change over time.

Estimation of the parameters are done by maximum likelihood. Some models can be evaluated using Newton-Raphson algorithms, but we rather use the NLMIXED procedure in SAS to find the estimates and their corresponding standard error numerically.

2. Minimum Bias and Classical Statistical Distributions

2.1 Minimum Bias

Historically, the minimum bias technique introduced by Bailey and Simon (1960) and Bailey (1963) were used to find the parameters of some classification rating systems. In these techniques, Bailey (1963) proposed to find parameters that minimise the biais of the premium by iterative algorithms. With the development of the Generalised Linear Models (GLM) (McCullagh and Nelder (1989)), actuaries now use statistical distributions to estimate the parameters of a rating system. It has been shown that the results obtained from this theory are very close to the ones obtained by the minimum bias technique (see Brown (1988), Mildenhall (1999), or Holler et al. (1999) from the 9^{th} exam of the Casualty Actuarial Society, that expose the link between the GLM and the minimum bias technique).

2.2 Poisson

From a statistical point of view, commonly, the starting point for the modeling of the number of reported claims $N_{i,t}$, for insured i at time t, is the Poisson distribution :

$$\Pr[N_{i,t} = n_{i,t}] = \frac{\lambda_{i,t}^{n_{i,t}} e^{-\lambda_{i,t}}}{n_{i,t}!}, \tag{2.1}$$

where the $\lambda_{i,t} = \exp(x'_{i,t}\beta)$. Because the Poisson distribution is a member of the exponential family, from which the GLM theory comes from, it has some useful property. For more informations about the technical part of the Poisson distribution for actuarial science, or even the GLM theory, we refer the interested reader to the guide of Feldblum and Brosius (2002).

However, it is well known that the Poisson distribution has some severe drawbacks, such as its equidispersion property, that limit its use (Gourieroux (1999)). To correct these disadvantages, the Negative Binomial distributions are common alternatives.

2.3 Negative Binomial

There are many ways to construct the Negative Binomial distribution, but the more intuitive one consists on the introduction of a random heterogeneity term θ of mean 1 and variance α in the mean parameter of the Poisson distribution. If the variable θ is following a Gamma distribution, having the follow density distribution :

$$f(\theta) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)}\theta^{1/\alpha-1}\exp(-\theta/\alpha), \tag{2.2}$$

well-known that this mixed model will result in a Negative Binomial distribution (NB) (Greenwood and Yule (1920) or Lawless (1987) and Dionne and Vanasse (1989) for an application with insurance data) :

$$\Pr[N_{i,t} = n_{i,t}] = \frac{\Gamma(n_{i,t}+\alpha^{-1})}{\Gamma(n_{i,t}+1)\Gamma(\alpha^{-1})}\Big(\frac{\lambda_{i,t}}{\alpha^{-1}+\lambda_{i,t}}\Big)^{n_{i,t}}\Big(\frac{\alpha^{-1}}{\alpha^{-1}+\lambda_{i,t}}\Big)^{\alpha^{-1}}, \tag{2.3}$$

where the $\lambda_{i,t} = \exp(x'_{i,t}\beta)$ and the function Γ is the gamma function, defined by $\Gamma(a) = \int_0^\infty e^{-t}t^{a-1}dt$. To ensure that the heterogeneity mean is equal to 1, both parameters of the Gamma distribution are chosen to be identical and equal to $1/\alpha$. From this, it can easily be proved that $E[N_{i,t}] = \lambda_{i,t}$ and $Var[N_{i,t}] = \lambda_{i,t} + \alpha\lambda_{i,t}^2$.

Cameron and Trivedi (1986) considered a more general class of Negative Binomial distributions having the same mean, but a variance of the form $\lambda_{i,t} + \alpha\lambda_{i,t}^p$. This kind of ditribution can be generated with an heterogeneity factor following a Gamma distribution with both parameters equal to $\lambda_{i,t}^{2-p}/\alpha$. A model with $p = 2$ is called $NB2$ distribution and is the previous one. When p is set to 1, it leads to the $NB1$ model :

$$\Pr[N_{i,t} = n_{i,t}] = \frac{\Gamma(n_{i,t}+\alpha^{-1}\lambda_{i,t})}{\Gamma(n_{i,t}+1)\Gamma(\alpha^{-1}\lambda_{i,t})}(1+\alpha)^{-\lambda_{i,t}/\alpha}(1+\alpha^{-1})^{-n_{i,t}}. \tag{2.4}$$

This model is interesting because its variance is equal to $\lambda_{i,t} + \alpha\lambda_{i,t} = \phi\lambda_{i,t}$, the same used in the Poisson GLM approach (such as the one used for the overdispersion correction of the Poisson distribution in the GENMOD procedure of SAS). We can note that the Poisson distribution is the limiting case of Negative Binomial distributions when the parameter α goes to zero.

parameter	Poisson		NB2		NB1	
b0	−2.6039	(0.0410)	−2.6046	(0.0434)	−2.6059	(0.0426)
b1	0.1047	(0.0337)	0.1059	(0.0356)	0.1167	(0.0348)
b2	−0.2106	(0.0390)	−0.2108	(0.0415)	−0.2134	(0.0404)
b3	−0.2721	(0.0361)	−0.2727	(0.0383)	−0.2803	(0.0373)
b4	0.1543	(0.0336)	0.1548	(0.0356)	0.1573	(0.0348)
b5	0.1373	(0.0291)	0.1383	(0.0306)	0.1441	(0.0303)
α	.	.	1.5744	(0.1095)	0.1102	(0.0077)
Log-Lik.	-27,133.8		-26,933.0		-26,926.4	

Tab. 2.1 – Time Independent Models

There are many possible distributions that can be used to model the heterogeneity, such as the Inverse-Gaussian heterogeneity (Holla (1966) or Willmot (1987), Dean et al. (1989) and Tremblay (1992) for an application with insurance data) that leads to a closed-form distribution. An other important heterogeneity distribution is the Log-Normal distributions (Hinde (1982)), but closed-form representation is impossible and numerical computations are needed (see Boucher et al. (2006c) for an application of all these models in insurance). Other Poisson mixture models are possible (see chapter 8 of Johnson et al. (1992)), but numerical techniques are also needed.

2.4 Numerical Application

Results of the application of these time independent models are shown in table 2.1. A comparison of the loglikelihoods exposes that the extra parameter of the Negative Binomial distributions improves the fit compared to the Poisson distribution. The estimations of the β parameters are approximately the same for all models, which is expected when the sufficient conditions for consistency are satisfied (see Gourieroux et al. (1984b) and Gourieroux et al. (1984a)).

3. Random Effects

3.1 Overview

To account for a dependence that can exist between all the contracts of the same insured, one of the most popular way is the use of a common individual term (Hausman et al. (1984)). Insurance data exhibits some variability that may be caused by the missing of some important classification

variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.). These hidden features are usually captured by the a individual random heterogeneity term.

Given the insured-specific heterogeneity term, θ_i, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. The joint probability function of $N_{i,1}, ..., N_{i,T}$ is thus given by

$$\begin{aligned} \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= \int_0^\infty \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\theta_i] g(\theta_i) d\theta_i \\ &= \int_0^\infty \left(\prod_{t=1}^{T} \Pr[N_{i,t} = n_{i,t}|\theta_i] \right) g(\theta_i) d\theta_i. \end{aligned} \tag{3.1}$$

Depending on the choices of the contidional distribution of the $N_{i,t}$ and the distribution distribution of θ_i, the model can exhibit particular properties as it seen below.

3.2 Endogeneous Regressors and Fixed Effects

In linear regression, correlation between the regressor and the error term leads to inconsistency of the estimated parameters (Mundlak (1978) or Hsiao (2003) for a review in case of longitundinal data). The same problem exists for the count data regression when $E[\theta|x_i] \neq E[\theta]$ (Mullahy (1997)) and it leads to biased estimates of parameters. In insurance, correlation between regressors and error term is often present (Boucher and Denuit (2006)) and may be caused by omitted variables that are correlated with the included ones.

As noted in Winkelmann (2003a), consistent estimates may be found if corrections are made to the standard estimation procedures. However, for insurance application, as shown by Boucher and Denuit (2006), standard methods of estimation, like classic maximum likelihood on joint distribution based on random effects, can still be used. Indeed, the resulting estimate of parameters, while being biased, represents the apparent effect on the frequency of claim, which is the interest when the correlated omitted variables cannot be used in classification.

3.3 Poisson Distribution

The simplest random effects model for count data is the Poisson distribution with an individual heterogeneity term that follows a specified distribution. Formally, we can express the classic Poisson random effects model as :

$$N_{i,t}|\theta_i \sim Poisson(\theta_i \lambda_{i,t}), \quad i = 1, ..., N \quad t = 1, ..., T,$$

where i represent an insured and t his covering period. As for the heterogeneous models of the cross-section data model seen at the begginning of this paper, many possible distribution for the random effects can be chosen. The use of a Gamma distribution can be used to express the joint distribution in a closed form. Indeed, the joint distribution of the number of claims, when the random effects follow a Gamma distribution of mean 1 and variance α, is equal to (Hausman et al. (1984)) :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] =$$
$$\left[\prod_{t=1}^{T} \frac{(\lambda_{i,t})^{n_{i,t}}}{n_{i,t}!}\right] \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + 1/\alpha)}{\Gamma(1/\alpha)} \left(\frac{1/\alpha}{\sum_{i=1}^{T} \lambda_{i,t} + 1/\alpha}\right)^{1/\alpha} (\sum_{t=1}^{T} \lambda_{i,t} + 1/\alpha)^{-\sum_{t=1}^{T} n_{i,t}}. \quad (3.2)$$

This distribution, which has been often applied (see chapter 36 of Johnson et al. (1996) for an overview), is known as a Multivariate Negative Binomial or Negative Multinomial. Note that this distribution can also be seen as the generalization of the bivariate Negative Binomial of Marshall and Olkin (1990). For this distribution, $E[N_{i,t}] = \lambda_{i,t}$ and $Var[N_{i,t}] = \lambda_{i,t} + \alpha\lambda_{i,t}^2$, so overdispersion can be accounted. Maximum likelihood estimates of the parameters and variance estimates for them are straightforward.

The generalization of the variance of the Multivariate Negative Binomial to obtain a form $\phi\lambda_{i,t}$, like the one obtained in the NB1 distribution cannot be done directly. Indeed, since the heterogeneity term needs to be individually specific and time invariant, introduction of covariates in the heterogeneity distribution cannot be allowed. Approximation of the covariates $x_{i,t}$ by $\hat{x}_i$ can be an interesting solution but cannot be used directly for prediction because characteristics of the risk at time $t + j$ is not known for $j > 0$.

Instead, *Pseudo*-Multinomial NB1 model can be used with the time invariant characteristics, like the sex of the insured, and can be modeled as $\gamma_i = \exp(y_i'\delta)$, where y_i is a subset of the $x_{i,t}$ that contains only these invariant covariates. Then, if we suppose that the heterogeneity is following a Gamma distribution with both parameters equal to γ_i^{1-k}/α, the joint distribution can be expressed as :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] =$$
$$\left[\prod_{t=1}^{T} \frac{(\lambda_{i,t})^{n_{i,t}}}{n_{i,t}!}\right] \frac{\Gamma(\sum_{i=1}^{T} n_{i,t} + \gamma_i^{1-k}/\alpha)}{\Gamma(\gamma_i^{1-k}/\alpha)} \left(\frac{\gamma_i^{1-k}/\alpha}{\sum_{i=1}^{T} \lambda_{i,t} + \gamma_i^{1-k}/\alpha}\right)^{\gamma_i^{1-k}/\alpha} (\sum_{t=1}^{T} \lambda_{i,t} + \gamma_i^{1-k}/\alpha)^{-\sum_{t=1}^{T} n_{i,t}}, \tag{3.3}$$

where $k = 1$ is the standard Negative Multinomial distribution (MVNB), $k = 0$ is the Pseudo Multinomial NB1 equivalence (PMVNB1). Expressed in its general form, the Poisson-Gamma distributions has the following moments :

$$\begin{aligned} E[N_{i,t}] &= E[E[N_{i,t}|\theta_i]] \\ &= \lambda_{i,t} E[\theta_i] \\ &= \lambda_{i,t}, \end{aligned} \tag{3.4}$$

$$\begin{aligned} Var[N_{i,t}] &= E[Var[N_{i,t}|\theta_i]] + Var[E[N_{i,t}|\theta_i]] \\ &= \lambda_{i,t} + \lambda_{i,t}^2 Var[\theta_i] \\ &= \lambda_{i,t} + \alpha(\lambda_{i,t}^2 \gamma_i^{k-1}), \end{aligned} \tag{3.5}$$

$$\begin{aligned} Cov[N_{i,t}, N_{i,t+j}] &= Cov[E[N_{i,t}|\theta_i], E[N_{i,t+j}|\theta_i]] + E[Cov[N_{i,t}, N_{i,t+j}|\theta_i]] \\ &= Cov[\lambda_{i,t}\theta_i, \lambda_{i,t+j}\theta_i] + 0 \\ &= \lambda_{i,t}\lambda_{i,t+j}\alpha\gamma^{k-1}, \quad j > 0. \end{aligned} \tag{3.6}$$

As for the cross-section data, other distributions can be chosen to model the random effects, such as the Inverse Gaussian or the LogNormal distributions (see Boucher and Denuit (2006) for an application in insurance), which result in distributions having the same two first moments as the MVNB, distinctions found using higher moments. All these random effects models come back to the Poisson distribution in the case where $\alpha \to 0$.

3.4 Negative Binomial

The Negative Binomial distribution can also be used with random effects, as shown by Hausman et al. (1984). Conditionally on the random effects δ_i, the authors used a NB1 identical to equation (2.4) except for a couple of parameter changes :

$$\Pr[N_{i,t} = n_{i,t}|\delta_i] = \frac{\Gamma(n_{i,t}+\lambda_{i,t})}{\Gamma(\lambda_{i,t})\Gamma(n_{i,t}+1)}\Big(\frac{\delta_i}{1+\delta_i}\Big)^{\lambda_{i,t}}\Big(\frac{1}{1+\delta_i}\Big)^{n_{i,t}}, \tag{3.7}$$

where $\lambda_{i,t} = \exp(x'_{i,t}\beta)$. Under this parametrization, the conditional distribution has the following moments :

$$\begin{aligned} E[N_{i,t}|\delta_i] &= \lambda_{i,t}/\delta_i, & (3.8)\\ Var[N_{i,t}|\delta_i] &= \lambda_{i,t}(1+\delta_i)/\delta_i^2 & (3.9)\\ &= E[N_{i,t}|\delta_i](1+\delta_i)/\delta_i. \end{aligned}$$

Thus, this conditional distribution implies overdispersion. Under the construction of (3.1), Hausman et al. (1984) assumed that the expression $\delta_i/(1+\delta_i)$ is following a Beta distribution with parameter (a,b), having mean $a/(a+b)$ and variance $ab/\big((a+b+1)(a+b)^2\big)$. Following the development of Hausman et al. (1984), the joint distribution can be expressed as :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \frac{\Gamma(a+b)\Gamma(a+\sum_t \lambda_{i,t})\Gamma(b+\sum_t n_{i,t})}{\Gamma(a)\Gamma(b)\Gamma(a+b+\sum_t \lambda_{i,t}+\sum_t n_{i,t})}\prod_t^T \frac{\Gamma(\lambda_{i,t}+n_{i,t})}{\Gamma(\lambda_{i,t})\Gamma(n_{i,t}+1)}. \tag{3.10}$$

As for the Poisson distribution with random effects, it is possible to use covariates to model a parameter of the Beta distribution, like $a = \exp(y'_i\beta)$. The moments of the Negative Binomial-Beta distribution are as follow :

$$\begin{aligned} E[N_{i,t}] &= E[E[N_{i,t}|\delta_i]]\\ &= \lambda_{i,t}E[1/\delta_i]\\ &= \lambda_{i,t}\frac{b}{a-1}, & (3.11) \end{aligned}$$

$$\begin{aligned} Var[N_{i,t}] &= E[Var[N_{i,t}|\delta_i]] + Var[E[N_{i,t}|\delta_i]] \\ &= \lambda_{i,t} E[(1+\delta_i)/\delta_i^2] + \lambda_{i,t}^2 Var[1/\delta_i] \\ &= \lambda_{i,t} \frac{(a+b-1)b}{(a-1)(a-2)} + \lambda_{i,t}^2 \left[\frac{(b+1)b}{(a-1)(a-2)} - \frac{b^2}{(a-1)^2} \right], \end{aligned} \tag{3.12}$$

$$\begin{aligned} Cov[N_{i,t}, N_{i,t+j}] &= Cov[E[N_{i,t}|\delta_i], E[N_{i,t+j}|\delta_i]] + E[Cov[N_{i,t}, N_{i,t+j}|\delta_i]] \\ &= Cov[\lambda_{i,t}/\delta_i, \lambda_{i,t+j}/\delta_i] + 0 \\ &= \lambda_{i,t}\lambda_{i,t+j} Var[1/\delta_i] \\ &= \lambda_{i,t}\lambda_{i,t+j} \frac{b}{a-1} \left(\frac{b+1}{a-2} - \frac{b}{a-1} \right), \quad j > 0. \end{aligned} \tag{3.13}$$

The distribution collapses to its Negative Binomial form for :

$$Var[\delta_i/(1+\delta_i)] = ab/\left((a+b+1)(a+b)\right) \to 0.$$

Other mixing distributions can be found using the negative binomial distribution, such as the negative binomial - triangular distribution proposed by Karlis and Xekalaki (2006). However, this distribution is not so interesting since its variance does not exist.

3.5 Predictive Distribution

The computations of the predictive distributions of panel data with random effects involve Bayesian analysis. Indeed, at each insured period, the random effects θ_i or δ_i can be updated for past claim experience, revealing some insured-specific informations. Formally :

$$\begin{aligned} \Pr[N_{i,T+1} = n_{i,T+1} | n_{i,1}, ..., n_{i,T}] &= \frac{\Pr(n_{i,1}, ..., n_{i,T+1})}{\Pr(n_{i,1}, ..., n_{i,T})} \\ &= \frac{\int \Pr(n_{i,1}, ..., n_{i,T+1}, \theta_i) d\theta_i}{\int \Pr(n_{i,1}, ..., n_{i,T}, \theta_i) d\theta_i} \\ &= \int \Pr(n_{i,T+1}|\theta_i) \left(\frac{\Pr(n_{i,1}, ..., n_{i,T}|\theta_i) g(\theta_i)}{\int \Pr(n_{i,1}, ..., n_{i,T}|\theta_i) g(\theta_i) d\theta_i} \right) d\theta_i \\ &= \int \Pr(n_{i,T+1}|\theta_i) \left(\frac{\left[\prod_t \Pr(n_{i,t}|\theta_i)\right] g(\theta_i)}{\int \left[\prod_t \Pr(n_{i,t}|\theta_i)\right] g(\theta_i) d\theta_i} \right) d\theta_i \\ &= \int \Pr(n_{i,T+1}|\theta_i) g(\theta_i | n_{i,1}, ..., n_{i,T}) d\theta_i, \end{aligned} \tag{3.14}$$

where $g(\theta_i|n_{i,1},...,n_{i,T})$ is the a posteriori distribution of the random effects θ_i. If this a posteriori distribution can be expressed in closed form, moments of the predictive distribution can be found easily be conditioning on the random effects θ_i.

Predictive and a posteriori distributions use Bayesian theory and are strongly related to credibility theory (Bühlmann (1967), Bühlmann and Straub (1970), Hachemeister (1975), Jewell (1975)). Linear credibility is a theory used to obtain premium based on a weighted average of past experience and a priori premium. An overview of the process of classification with random effects and predictive distributions can be found in Buhlmann and Gisler (2005).

3.5.1 Poisson

As it is well known, we found that the a posteriori distribution of the heterogeneity term for the Poisson model with Gamma random effects is also a Gamma distributed with parameters $\sum_t \lambda_{i,t} + \gamma_i^{(1-k)}/\alpha$ and $\sum_t n_{i,t} + \gamma_i^{(1-k)}/\alpha$. We also see that the generalisation into Pseudo-MVNB1 can be interesting for the prediction of premium. The future premium (frequency part), which is equal to the expected number of reported claims, is equal to :

$$E[N_{i,t+1}|N_{i,1},...,N_{i,t}] = \lambda_{i,t+1}\frac{\sum_t n_{i,t} + \gamma_i^{(1-k)}/\alpha}{\sum_t \lambda_{i,t} + \gamma_i^{(1-k)}/\alpha}. \quad (3.15)$$

The variance of the heterogeneity distribution depends on time-invariant characteristics of the insureds. Then, the impact on premium of having a claim is more important for low variance random effects and consequently, premium modifications for negative components of δ are more severe.

3.5.2 Negative Binomial

The *a posteriori* density of the heterogeneity term of the Negative Binomial with Beta random effects, proposed by Hausman et al. (1984), has also a closed form. Indeed, using equation (3.14), it can be shown that the ratio $\delta_i/(1+\delta_i)$ follows a Beta distribution with parameters $\sum_t \lambda_{i,t} + a$ and $\sum_t n_{i,t} + b$. Consequently, for this model, the frequency part of the future premium can be expressed as :

parameter	MVNB		Pseudo-MVNB1		NB-Beta	
b0	−2.6241	(0.0447)	−2.6257	(0.0447)	0.8198	(0.1522)
b1	0.1080	(0.0407)	0.1184	(0.0385)	0.1170	(0.0404)
b2	−0.2012	(0.0398)	−0.2013	(0.0398)	−0.2063	(0.0405)
b3	−0.2493	(0.0385)	−0.2493	(0.0385)	−0.2600	(0.0390)
b4	0.1453	(0.0375)	0.1445	(0.0375)	0.1499	(0.0378)
b5	0.1425	(0.0332)	0.1429	(0.0332)	0.1479	(0.0334)
α	0.8798	(0.0431)	0.0648	(0.0042)	.	.
a	.	.	.	.	43.2890	(5.6332)
b	.	.	.	.	1.3529	(0.0802)
Log-Lik.	-26,690.9		-26,689.8		-26,657.4	

Tab. 3.1 – Random Effects Models

$$\begin{aligned} E[N_{i,t+1}|N_{i,1},...,N_{i,t}] &= E[E[N_{i,t+1}|N_{i,1},...,N_{i,t},\delta_i]] \\ &= \lambda_{i,t}E[1/\delta_i|N_{i,1},...,N_{i,t}] \\ &= \lambda_{i,t+1}\frac{\sum_t n_{i,t}+b}{\sum_t \lambda_{i,t}+a-1}, \end{aligned} \tag{3.16}$$

which has the same form as the future premium with the Poisson-Gamma model, but allows more flexibility since an additional parameter is used to calculate the premium. A more formal comparison between future premiums and predictive distributions of the models is made in section 8.

3.6 Numerical Application

The models with random effects have been applied and the results are shown in table 3.1. If one looks at the loglikelihood, we can note that the models with time dependence exhibit a better fit than the one with no dependence between contracts of the same insured. Note also that the model based on the Negative Binomial distribution has a lower loglikelihood value that the others.

The pseudo MVNB1 model includes the parameter $b1$, representing the sex of the driver that does not change in time, in the gamma component, such as $\gamma_i = \exp(b_0 + b_1 x_{i1})$. Parameter esti-

mates are approximately equal to each other, but they do exhibit small differences from the time independent models. Indeed, such differences are expected since individual insured characteristics are captured by the random effect. The p-values of the extra parameters that account for the dependence between contracts of the same insureds indicate the significance of the random effects, but note that precautions must be taken in this analysis, as we will see later.

To be compared to the other models, the intercept of the NB-Beta distribution must be adapted. Indeed, the intercept must be modified to include the ratio $\frac{b}{a-1}$, as shown by equation (3.11). Using this modification, the *ajusted* intercept is equal to -2.62255, which is very close to the intercepts of the Multinomial NB models.

4. Past experience

4.1 Overview

To account for the time dependence, an intuitive approach can be the introduction of the past experience of the insured as covariate (Gerber and Jones (1975), Sundt (1988)). This model interprets past claims as a factor that changes the mean of the distribution. The conditional distribution uses a Markovian property where only the number of claims of the last period is considered. It would result in Poisson or Negative Binomial distributions with mean equal to $\exp(x_{i,t}\beta + c\ n_{i,t-1})$. The variable c can also be computed by covariates, such as $c_i = y_i'\gamma$, where a negative (positive) value of c implied negative (positive) dependence.

However, as stated in Gourieroux and Jasiak (2004), the stationarity properties of this model cannot be established and premiums for new insureds cannot be computed. Thus, contracts that come from the first year of the data cannot be used, which is clearly not satisfactoring. Neverthless, this model can have some uses, as proposed by Cameron and Trivedi (1998), since this regression of n_t on n_{t-1} can be used to test the time dependence by a standard t test on the parameter.

4.2 Conditional Distributions with Artificial Marginal Distribution

4.2.1 Poisson

This kind of model is closely related to a model described in Berkhout and Plug (2004), where n_1 is Poisson distributed with parameter $\lambda_1 = \exp(x_1\beta)$, while $n_t|n_{t-1}$, $t = 2, ..., T$ are Poisson with mean $\lambda_t = \exp(x_t'\beta + \phi n_{t-1})$. Then, conditional expected value and conditional variance are both equal to λ_t. The joint distribution of all contracts of insured i for this model, that we can call

Multivariate Poisson with Artificial Marginal Distribution (MPAM), can be expressed as (subscript i removed for simplicity) :

$$\begin{aligned} \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= g(n_1)g(n_2|n_1)...g(n_T|n_{T-1}) \\ &= \prod_{t=1}^{T} \frac{\lambda_t^{n_t} e^{-\lambda_t}}{n_t!}, \end{aligned} \tag{4.1}$$

where even ϕ can be modeled with covariates, such as $\phi_t = y_t'\gamma$. The distribution exhibits autoregressive dependence since the value of n_t depends only on n_{t-1}. Though it has no real impact for insurance applications, note that closed form for the marginal distribution of N_t, $t = 2, ..., T$ cannot be found. However, factorial moments of the joint distribution can be defined. As proved by Berkhout and Plug (2004), the $(r, s)^{th}$ factorial moment of N_{t-1} and N_t are equal to :

$$\begin{aligned} E[N_t(N_t - 1)...(N_t - r + 1)N_{t-1}(N_{t-1} - 1)...(N_{t-1} - s + 1)] = \\ \lambda_{t-1}^s e^{\lambda_{t-1}(\exp(r\phi)-1)+rx'\beta+rs\phi}. \end{aligned} \tag{4.2}$$

Using this last equation, the mean and the variance of the marginal distribution can be computed :

$$E[N_t] = e^{\lambda_{t-1}(\exp(\phi_{t-1})-1)}\lambda_t \tag{4.3}$$

$$Var[N_t] = E[N_t] + E[N_t]^2(e^{\lambda_{t-1}(\exp(\phi_{t-1})-1)^2} - 1), \tag{4.4}$$

where we can see that marginal distributions imply overdispersion for all ϕ different than 0. The covariance between two successive events can also be calculated :

$$Cov(N_{t-1}, N_t) = \lambda_{t-1}E[N_t](\exp(\phi_{t-1}) - 1). \tag{4.5}$$

Obviously, when the parameter ϕ is set to 0, the model collapses to the product of independent Poisson distributions.

4.2.2 **Negative Binomial**

Using the same development as for the MPAM distribution, we can also use conditional Negative Binomial distributions to construct a model that accounts for past experience. Here, we choose to

use the NB1 distribution, but it is also possible to use other forms of Negative Distributions. The Multivariate Negative Binomial with Artificial Marginal Distribution (MNBAM), also shows a conditional expected value of $\lambda_{i,t}$, while its conditional variance is equal to $\lambda_{i,t} + \alpha\lambda_{i,t}$. Using equation (4.1), the joint distribution can be expressed as :

$$\begin{aligned} \Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] &= g(n_{i,1})g(n_{i,2}|n_{i,1})...g(n_{i,T}|n_{i,T-1}) \\ &= \prod_{t=1}^{T} \frac{\Gamma(n_{i,t} + \alpha^{-1}\lambda_{i,t})}{\Gamma(n_{i,t}+1)\Gamma(\alpha^{-1}\lambda_{i,t})}(1+\alpha)^{-\lambda_{i,t}/\alpha}(1+\alpha^{-1})^{-n_{i,t}}, \quad (4.6) \end{aligned}$$

with, as for the MPAM model, $\lambda_{i,1} = \exp(x_{i,1}\beta)$, while $\lambda_{i,t} = \exp(x'_{i,t}\beta + \phi n_{i,t-1})$, for $t = 2, 3, ...$ The marginal properties of contracts $2, 3, ..., T$ are useless for insurance application since new insureds are classified by $g(n_{i,1})$ and the distributions of other insureds are only interesting for conditional analysis on past experience. But, once again, note that no closed-form of the marginal distribution for contracts $2, 3, ..., T$ can be found and even the method used by Berkhout and Plug (2004) cannot be helpful for the finding of factorial moments. However, by iterative conditioning on each heterogeneity term, moments can be found. When the parameter $\alpha \to 0$, the model becomes the MPAM. Additionaly, the model collapses to the product of independent NB distributions if the parameter ϕ is set to 0.

4.3 Predictive Distribution

The MPAM model of Berkhout and Plug (2004) and its generalisation into NB distributions have a Markovian property and only depend on the number of reported claims of the last insured period. Generalization into models that incorporate higher order dependences could be interesting. Consequently, the predictive distributions of this kind of model, conditionaly on a past insured period, are standard Poisson and Negative Binomial distributions.

4.4 Numerical Application

The estimated parameters of the Multivariate Poisson and Negative Binomial with Artificial Marginal distributions are shown in table 4.1. It shows that $b5$, representing the power of the vehicle, has an impact on the dependence between the past number of claims at time $t-1$. Indeed, vehicle with power larger or equal to 5500 cc exhibit less time dependence. For the other parameters, estimated values are quite the same as those evaluated by the time independent models. Because the Poisson marginal distribution implies marginal equidispersion, it is not surprising to see that model based on an NB1 marginal distributions exhibit a better fit.

parameter		MPAM		MNB1AM	
λ :	b0	−2.6549	(0.0417)	−2.6582	(0.0433)
	b1	0.0995	(0.0337)	0.1103	(0.0348)
	b2	−0.2152	(0.0390)	−0.2184	(0.0404)
	b3	−0.2807	(0.0361)	−0.2882	(0.0373)
	b4	0.1440	(0.0336)	0.1466	(0.0347)
	b5	0.1490	(0.0305)	0.1575	(0.0316)
ϕ :	c0	−0.3023	(0.0877)	−0.2886	(0.0892)
	c5	−0.2609	(0.1076)	−0.2804	(0.1106)
α		.	.	0.1064	(0.0076)
Log-Lik.		-26,985.0		-26,790.7	

TAB. 4.1 – Artificial Marginal Distributions

5. Integer Valued Autoregression Models

5.1 Overview

An other method to include the time dependence in the classification process is the integer valued autoregressive process, due to Al-Osh and Alzaid (1987) and McKenzie (1988) (see Gourieroux and Jasiak (2004) for an overview in an insurance context). As linear model, the autoregressive integer process can be of any order, but for simplicity, we restrict ourselves to AR(1) model. The Integer Valued Autoregression Models of order 1, noted INAR(1) is defined by the recursive equation :

$$n_{i,t} = \rho \circ n_{i,t-1} + I_{i,t}, \tag{5.1}$$

where $\circ$, the binomial thinning operator (Steutel and Harn (1979)), is defined as $\rho \circ n = \sum_{i=1}^{n} Y_i = B(\rho, n)$, where Y_i is a sequence of binary random variable Y_i, independent from n, where $\Pr[Y_i = 1] = 1 - \Pr[Y_i = 0] = \rho$, $\rho \in [0, 1]$. The random variable $I_{i,t}$ is i.i.d. and independent from $n_{i,t-1}$ with discrete non-negative values.

The INAR(1) distribution models the observation throught two different components that facilitate the marginal analyses. As expressed in Freeland and McCabe (2004), this distributions can be interpreted as a birth and death process, where the component $\rho \circ n_{i,t-1}$ represents the survivors

of the past period and $I_{i,t}$ the birth component. For insurance applications, we can interpret the parameters differently. Indeed, $\rho \circ n_{i,t-1}$ can be roughly seen as the number of claims due to the impact of the insured's past claims on his driving behavior, while $I_{i,t}$ would represent his number of claims caused by random variation.

The INAR(1) model has Markovian property :

$$\Pr(n_{i,t}|n_{i,t-1}, n_{i,t-2}, ...) = \Pr(n_{i,t}|n_{i,t-1}), \tag{5.2}$$

This property facilitates the expression of the joint density of $n_{i,1}, n_{i,2}, ..., n_{i,T}$:

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \Pr(N_{i,T} = n_{i,T}|n_{i,T-1}) ... \Pr(N_{i,2} = n_{i,2}|n_{i,1}) \Pr(N_{i,1} = n_{i,1}), \tag{5.3}$$

where the distribution of $f(n_{i,1})$ must be chosen to obtain a stationnary INAR(1) process, where the marginal distribution of each $n_{i,t}$ can be computed easily. This choice is obtained by a check of the stationarity of the process :

$$E[N_{i,t}] = \frac{E[I_{i,t}]}{1-\rho}, \tag{5.4}$$

$$Var[N_{i,t}] = \frac{\rho E[I_{i,t}] + Var[I_{i,t}]}{1-\rho^2}, \tag{5.5}$$

where the results come from conditional computations and the fact that the expected value and the variance of $n_{i,t}$ and $n_{i,t-1}$ must be equal to ensure stationarity.

As shown in Brännas (1995), the parameters can be modeled with covariates, such as $\lambda_{i,t} = \exp(x_{i,t}\beta)$ and $\rho_{i,t} = logit(z_{i,t}\gamma) = \exp(z_{i,t}\gamma)/(1 + \exp(z_{i,t}\gamma))$ Under this parametrisation, the two first moments of this process are expressed as :

$$E[N_{i,t}] = \rho_{i,t} E[n_{i,t-1}] + E[I_{i,t}], \tag{5.6}$$

$$Var[N_{i,t}] = \rho_{i,t}^2 Var[n_{i,t-1}] + \rho_{i,t}(1-\rho_{i,t}) E[n_{i,t-1}] \lambda_{i,t} + Var[I_{i,t}]. \tag{5.7}$$

Conditionaly on a time horizon of lenght h, covariance between these two events is expressed as :

$$Cov[N_{i,t}, N_{i,t-h}] = \left[\prod_{j=0}^{h-1} \rho_{i,t-j}\right] Var[N_{i,t-h}]. \tag{5.8}$$

5.2 Poisson

The Poisson-INAR(1) is obtained when the random variable $I_{i,t}$ is following a Poisson distribution of mean $\lambda_{i,t}$. The Poisson distribution can be chosen since the convolution product of two Poisson distributions is still a Poisson distribution. This characterization of the INAR process is the most widely used in pratice. Under this hypothesis, a stationnary INAR(1) process is obtained, where :

$$\Pr(N_{i,1} = n_{i,1}) = \exp\left[\frac{-\lambda_{i,1}}{1-\rho}\right] \frac{[\lambda_{i,1}/(1-\rho)]^{n_{i,1}}}{n_{i,1}!}, \tag{5.9}$$

$$\Pr(N_{i,t} = n_{i,t} | n_{i,t-1}) = \sum_{j=0}^{\min(n_{i,t}, n_{i,t-1})} \binom{n_{i,t-1}}{j} \rho^j (1-\rho)^{n_{i,t-1}-j} e^{-\lambda_{i,t}} \frac{\lambda_{i,t}^{n_{i,t}-j}}{(n_{i,t}-j)!}, \tag{5.10}$$

where the number of reported claims $n_{i,t}$ has a marginal Poisson distribution. Under this hypothesis, conditionaly on a time horizon of lenght h, the two first moments of this process are expressed as :

$$E[n_{i,t}|n_{i,t-h}] = \rho^h n_{i,t-h} + \frac{1-\rho^h}{1-\rho}\lambda_{i,t} \tag{5.11}$$

$$\begin{aligned} Var[n_{i,t}|n_{i,t-h}] &= \rho^h(1-\rho^h)n_{i,t-h} + \frac{1-\rho^h}{1-\rho}\lambda_{i,t} \\ &= E[n_{i,t}|n_{i,t-h}] - \rho^{2h} n_{i,t-h}, \end{aligned} \tag{5.12}$$

From the last equation, we see that the Poisson-INAR(1) distribution implies underdispersion, which is not desirable for insurance data. For $\rho = 0$, the distribution collapses to a standard Poisson distribution.

5.3 Negative Binomial

In order to account for the overdispersion of the data, the first-order INAR time series model, with Negative Binomial marginal distributions (McKenzie (1986) and Al-Osh and Alzaid (1993) and Bockenholt (1999),Bockenholt (2003)) can be an interesting alternative. The convolution product of two Negative Binomial distributions is again Negative Binomial for the NB1 form (see equation (2.4)) when the two distributions share in common the parameter α. Indeed, for two independent variables i having NB1(λ_i, α) distribution, the sum is also NB1 with mean $\lambda_1 + \lambda_2$ and variance $(\lambda_1+\lambda_2)(1+\alpha)$ (see Winkelmann (2003a) for a formal proof using probability generating functions).

Using $I_{i,t} \sim NB1(\lambda_{i,t}, \alpha)$, the marginal distribution of $n_{i,1}$ must be chosen to obtain a stationnary process. Using equations (5.4) and (5.5), parameters of the marginal distribution of $n_{i,t}$ are $\lambda_{i,t}/(1-\rho)$ and $\frac{\rho+\rho^2+\alpha}{1-\rho^2}$.

$$\Pr(N_{i,1} = n_{i,1}) = \frac{\Gamma(n_{i,1} + \lambda_{i,1}/(\alpha(1-\rho)))}{\Gamma(n_{i,1}+1)\Gamma(\lambda_{i,t}/(\alpha(1-\rho)))}(1+\alpha)^{-\lambda_{i,t}/(\alpha(1-\rho))}(1+\alpha^{-1})^{-n_1} \tag{5.13}$$

$$\Pr(N_{i,t} = n_{i,t}|n_{i,t-1}) = \sum_{j=0}^{\min(n_{i,t},n_{i,t-1})} C^j_{n_{i,t-1}}\rho^j(1-\rho)^{n_{i,t-1}-j}\frac{\Gamma(n_{i,t}-j+\alpha^{-1}\lambda)}{\Gamma(n_{i,t}-j+1)\Gamma(\alpha^{-1}\lambda_{i,t})}(1+\alpha)^{-\lambda_{i,t}/\alpha}(1+\alpha^{-1})^{-n_{i,t}+j}, \tag{5.14}$$

where the number of reported claims $n_{i,t}$ has a marginal Poisson distribution. Under this hypothesis, conditionnaly on a time horizon of lenght h, the two first moments of this process are expressed as :

$$E[N_{i,t}|n_{i,t-h}] = \rho^h n_{i,t-h} + \frac{1-\rho^h}{1-\rho}\lambda_{i,t} \tag{5.15}$$

$$\begin{aligned} Var[N_{i,t}|n_{i,t-h}] &= \rho^h(1-\rho^h)n_{i,t-h} + \frac{1-\rho^h}{1-\rho}\lambda + \frac{1-\rho^{2h}}{1-\rho^2}\lambda\alpha \\ &= E[N_{i,t}|n_{i,t-h}] - \rho^{2h}n_{i,t-h} + \frac{1-\rho^{2h}}{1-\rho^2}\lambda\alpha. \end{aligned} \tag{5.16}$$

For large value of h, conditional distribution allows for overdispersion, while small values of h allow overdispersion depending on values of ρ and α. For null value of the parameter ρ, the NB1-INAR(1) model is the NB1 distribution, while in the situation where $\alpha \to 0$, the NB1-INAR(1) collapses to the Poisson-INAR(1) distribution.

5.4 Predictive Distribution

As for the Conditional Distributions with Artificial Marginal distributions seen in past section, the INAR(1) models also share the Markovian property. Using generalisation into higher INAR models, it is possible to extend the dependence into more insured periods. The INAR models are directly constructed to be used as predictive distributions.

5.5 Numerical Application

We applied our insurance data to the INAR models seen above and the estimated parameters are shown in table 5.1. As the MPAM and the MNB1AM models, the covariate related to the vehicle power ($b5$) is also significant in the parameter involving time dependence. However, even if it is interesting at this point, it is important to note that this parameter in the INAR models does not represent the same thing as the one in the Artifical Margin models. Consequently, it must be interpreted differently. Other estimated parameters of the covariates are, once again, very similar to those obtained by the time independent models.

Model based on NB1 distribution exhibits a better fit to the one based on the Poisson distribution when we observe the loglikelihood. Like the MPAM and MNB1AM distributions, the marginal equidispersion can account this observation but also important is the conditional equidispersion implied by the Poisson-INAR model.

Neverthless, the fits of the INAR models are disappointing. Indeed, the loglikelihoods obtained from these models are much smaller that the ones obtained by the random effects or the Artificial Marginal distributions. It can be caused by the lack of flexibility of the model. Indeed, the time dependence is too much limited by the number of claims reported in the last insured period. Consequently, for example, no time dependence is implied by an insured without claim in his last contract.

	Poisson				NB1			
parameter	β		γ		β		γ	
b0	−2.6859	(0.0429)	2.5270	(0.1421)	−2.6824	(0.0445)	2.5828	(0.1484)
b1	0.1022	(0.0351)	.	.	0.1110	(0.0362)	.	.
b2	−0.2066	(0.0403)	.	.	−0.2125	(0.0416)	.	.
b3	−0.2652	(0.0372)	.	.	−0.2749	(0.0385)	.	.
b4	0.1499	(0.0349)	.	.	0.1519	(0.0361)	.	.
b5	0.1616	(0.0310)	0.3649	(0.1701)	0.1710	(0.0322)	0.4192	(0.1814)
α	.	.	.	.	0.1147	(0.0081)	.	.
Log-Lik.	-26,993.6				-26,816.0			

Tab. 5.1 – INAR Models

6. Common Shock Models

6.1 Overview

Common Shock models, where marginal distributions are standard count distributions could seem interesting to model all contracts of an individual. This distribution is based on a convolution structure. The dependence between contracts of the same insured comes from a common individual random variable that is added to each time period (Holgate (1964) for the bivariate case). Suppose $T+1$ random variables $M_1, M_2, ..., M_T$ and U with respective parameters $\lambda_{i,1}, \lambda_{i,2}, ..., \lambda_{i,T}$ and μ_i. With the assumption that the random variables M_t and U are independent, the common schock model is defined as the T-joint distribution of $M_{i,1}+U_i, M_{i,2}+U_i, ..., M_{i,T}+U_i$. Using appropriate count distributions, these models show interesting properties and can be easily tractable. This model can be interpreted as if an individual specificity of an insured affects all his contracts. It is closely related to the random effects models, but instead of having an impact on the mean parameter of all the marginal distributions, the model impacts equally the number of reported claims of all contracts.

6.2 Poisson

Multivariate Poisson distribution (MVP) is a common shock model where all random variables $M_1, M_2, ..., M_T$ and U are Poisson distributed. Using the property of the sum of independent Poisson variables, each component of this distribution has Poisson marginal distribution with mean and

variance equal to $\lambda_{i,t}+\mu_i$. Note that the common shock can also be modeled as $\mu_i = \exp(y_i'\gamma)$, using time-invariant covariates such as the sex of the driver in our data, which allowed the dependence between contracts to be different for men or women drivers for example.

The covariance between contracts of the same insured, say contracts t and $t+j$, can easily be computed and is equal to μ_i, the mean shock variable. Since U is Poisson distributed, the model excludes zero and negative correlation. Moreover, it also excludes correlation values higher than $\mu_i/(\mu_i + \min(\lambda_t, \lambda_{t+j}))$.

The joint probability function of the MVP distribution for insured i is expressed as :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \exp[-(\mu_i + \sum_t \lambda_{i,t})] \sum_{j=0}^{s_i} \frac{\mu_i^j}{j!} \frac{\lambda_{i,t}^{n_{i,t}-j}}{(n_{i,1}-j)!} ... \frac{\lambda_{i,T}^{n_{i,T}-j}}{(n_{i,T}-j)!}, \tag{6.1}$$

where $s_i = \min(n_{i,1}, ..., n_{i,T})$. From the joint probability, we can see that the common shock cannot exceed any of the $n_{i,t}$. Conditional distribution of $n_{i,t}$ given $n_{i,t-1}$ can be computed without problem since it involves ratios of Poisson distribution that are tractable easily. Interested readers can refer to Kocherlatoka and Kocherlatoka (1992) for details about the MVP distribution. The MVP model collapses to the Poisson distribution for $\mu \to 0$.

6.3 Negative Binomial

By the same shock process, Winkelmann (2000) constructed a common shock model for the Negative Binomial distribution that we called common shock Negative Binomial (CSNB) [1]. As it involves sums of distributions, the CSNB distribution uses NB1 distribution, as for the integer value autoregressive models seen earlier. On this hypothesis, the marginal distribution of each contract is NB1 distributed with mean $\lambda_{i,t} + \mu_i$ and variance $(\lambda_{i,t} + \mu_i)(1 + \alpha)$, which allows overdispersion.

The covariance implied between contracts t and $t+j$ of the same insured can be shown to be equal to $\mu_i(1+\alpha)$, while the correlation is exclusively positive with values less than $\mu_i/(\mu_i+\min(\lambda_t, \lambda_{t+j}))$.

As for the MVP distribution, the CMNB is expressed by convolution :

[1] We do not call this model Multivariate Negative Binomial or Negative Multinomial distributions, since these names refer to Poisson distributions sharing the same heterogeneity variable, as seen in section 3.

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \sum_{j=0}^{s_i} f_{NB}(j) \prod_{k=1}^{T} f_{NB}(n_{i,k} - j), \tag{6.2}$$

where $s_i = \min(n_{i,1}, ..., n_{i,T})$ and $f_{NB}(n_{i,t})$ is expressed with equation (2.4). An overview of multivariate distributions for count data is presented in Winkelmann (2003a). As for the MVP distribution, the CSNB model collapses to the NB1 distribution for $\mu \to 0$. For $\alpha \to 0$, the CSNB is not different from the MVP distribution.

6.4 Covariance Structure

Under these multivariate distribution, the dependence between each contracts of the same insured is the same. Using a combination of common shock variables, Karlis and Meligkotsidou (2005) proposed a multivariate Poisson model with larger covariance structure, where the generalisation into common shock Negative Binomial model can easily be done.

6.5 Predictive Distribution

For the multivariate Poisson and the multivariate Negative Binomial, the predictive distribution can be computed easily. Supposing that the common shock also affects the number of reported claims at time $T + 1$, the conditional distribution can be evaluated by Bayesian analysis, such as :

$$\Pr(N_{i,T+1} = n_{i,T+1} | n_{i,1}, ..., n_{i,T}) \quad = \quad \frac{f(n_{i,1}, ..., n_{i,T}, n_{i,T+1})}{f(n_{i,1}, ..., n_{i,T})}, \tag{6.3}$$

by using equations (6.1) or (6.2). For bivariate distributions, it can be shown that the a posteriori distribution of the common shock variable is following a binomial distribution, allowing a linear regression relationship between n_T and n_{T+1}. However, for multivariate variables, the model cannot be expressed so easily and numerical computation, or simulation techniques such as Markov Chain Monte Carlo (MCMC) can be helpful.

6.6 Numerical Application

As seen in equations (6.1) and (6.2), the common shock in both multivariate models cannot exceed any of the $n_{i,t}$. Consequently, the multivariate distributions cannot be used for modeling the number of reported claims in insurance. Indeed, in our data and in practically all automobile insurance portfolio, a great proportion of insureds does not report a single claim. Then, $s_i = \min(n_{i,1}, ..., n_{i,T}) = 0$ which results in a null μ_i, where the joint distribution becomes a product of

independent Poisson or NB1 distribution. Indeed, for illustration, the maximum correlation between two contracts of insureds who did not report a claim, equal to $\mu_i/(\mu_i + \min(\lambda_t, \lambda_{t+j}))$, is 0.

7. Copula Approach

7.1 Overview

An other possibility to model dependences between contracts of the same insured is the use of the copulas theory. Several papers used copulas to model joint distribution between discrete random variables, such as Lee (1983) and van Ophem (1999) who used a Gaussian copula, Vandenhende and Lambert (2000) who used Archimedean copulas and Cameron et al. (2004) who used both Gaussian and Archimedian copulas. This kind of distribution can be very interesting since it models the dependence of the marginal distribution, which is very different from the dependence constructed by the other models seen so far.

A copula is defined as a multivariate distribution whose one-dimensional margins are uniform on [0,1]. Then, $C(u_1, ..., u_T) = P(U_1 \leq u_1, .., U_T \leq u_T)$ is a function defined on [0,1] to [0,1]. Copulas allow creation of multivariate distribution with defined marginal distributions. More specifically, if C is a copula and $F_{X_1}, ..., F_{X_T}$ are marginal distribution, then $C(F_{X_1}(x_1), ..., F_{X_T}(x_T))$ is a multivariate distribution. See Nelsen (1999) or Joe (1997) for more details about copulas, or Frees and Valdez (1999) for an overview of possible uses in actuarial science.

A fundamental theorem in copulas theory has been given by Sklar (1959) that shows that any joint distribution $H(x_1, ..., x_T)$ with marginals $F_{X_1}, ..., F_{X_T}$ can be written with copula, such as $H(x_1, ..., x_T) = C(F_{X_1}(x_1), ..., F_{X_T}(x_T))$. Additionnaly, when the random variables $X_1, ...X_T$ are continuous, it can be shown that C is unique. This conclusion does not apply for discrete data. However, as shown by Nelsen (1999), it does not create problem since there exists a unique subcopula such that the construction of joint distribution is only determined on the range of margins.

7.2 Archimedean Copulas

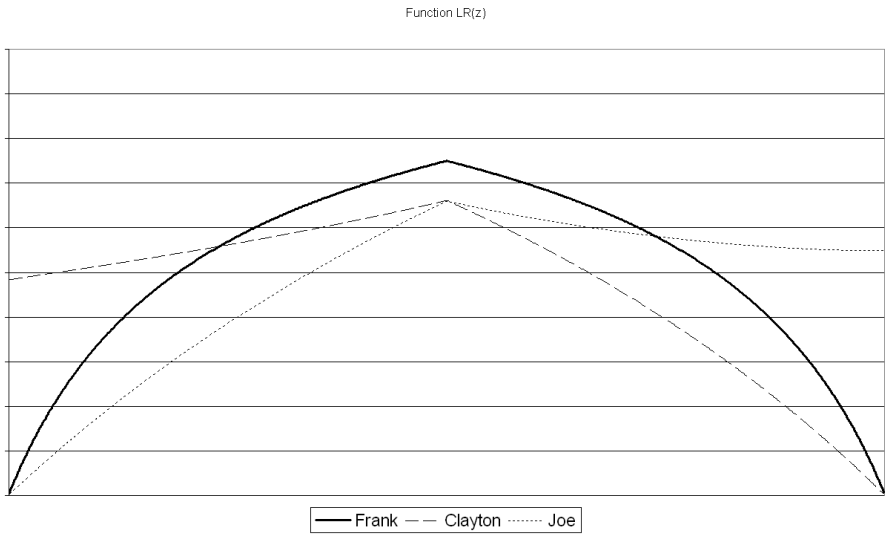
Numerous kind of copulas can be used to model dependence between random variables. However, we work with one of the most popular family of copulas, called Archimedean copulas, that can be used easily to model dependence between more than 2 variables. These copulas are created from a particular generating function ϕ, such as $C(u_1, ..., u_T) = \phi^{-1}[\phi(u_1) + ... + \phi(u_T)]$, where $U_1, ..., U_T$ are a series of random variables with uniform margins on [0,1].

For $T = 2$, the generating function $\phi : [0, 1] \to [0, \infty)$ is continuous and strictly decreasing such that $\phi(0) = \infty$ and $\phi(1) = 0$. For $T > 2$, another requirement needs ϕ^{-1} to be completely monotonic on $[0, \infty)$ (Kimberling (1974)). (see Nelsen (1999) for more details about different families of copulas).

In this paper, we use 3 differents Archimedean copulas with one parameter. There are many ways to visualize copulas, but we chose to use the Tail Concentration Functions (Joe (1997)) to focus of the dependence that can imply each copulas. The Tail Concentration Function $LR(z)$ is contructed on left (small values) and right (large values) dependencies. Formally, with a simplification into 2 dimensions, say random variables U and V, this function can be expressed as :

$$LR(z) = \begin{cases} L(z) & \text{if } 0 \leq z < 0.5 \\ R(z) & \text{if } 0.5 \leq z \leq 1 \end{cases}, \tag{7.1}$$

where the L(z) function is the probability of $U < z$ conditionnaly on $V < z$, while R(z) is the probability of $U > z$ conditionnaly on $V > z$. For the three copulas studied in this paper, the Frank, the Clayton and the Joe copulas, each LR(z) is shown in the following graphic :



The Frank's copula is a very popular Archimedean copula that exhibits radial symetry, without implying tail dependence. The copula can be expressed as :

$$C_\theta(u_1, ..., u_T) = -\frac{1}{\theta} \log \left[1 + \frac{\prod_{i=1}^{T} (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{T-1}} \right]. \tag{7.2}$$

Because copulas involving more than two random variables can only imply positive dependence, the parameter θ of the Frank's copula is exclusively positive. When the parameter tends toward zero, the Frank's copula becomes the independance copula (Π), expressed as the product of each marginal distributions. When the parameter is going to infinity, the Frank's copula reaches the upper Frechet bound (M).

As seen in the LR(z) function, the Clayton's copula shows left-tail dependence, implying that small value of one variable brings high probability of having small value for the other one. The T-variate Clayton's copula is expressed as :

$$C_\theta(u_1, ..., u_T) = \left[u_1^{-\theta} + ... + u_T^{-\theta} - T + 1 \right]^{-1/\theta}, \tag{7.3}$$

As for the Frank's copula, parameter of the Clayton's copula is strictly positive, Π copula is reach for $\theta \to 0$ and M copula for $\theta \to \infty$.

At last, the Joe's copula shows right dependence and is expressed as :

$$C_\theta(u_1, ..., u_T) = 1 - \left[1 - \prod_{i=1}^{T} \left(1 - (1 - u_i)^\theta \right) \right]^{1/\theta}, \tag{7.4}$$

where the parameter $\theta \leq 1$, Π copula reached for $\theta = 1$ and M copula for $\theta \to \infty$.

7.3 Dependence

7.3.1 Exchangeable Dependence

Supposing a joint distribution modeled by marginal distributions Poisson or Negative Binomial, we can suppose an exchangeable dependence based on :

$$\Pr[N_{i,1} = n_{i,1}] = \Pr[N_{i,1} \leq n_{i,1}] - \Pr[N_{i,1} \leq n_{i,1} - 1] \tag{7.5}$$

and

$$\begin{aligned}\Pr[N_{i,1}=n_{i,1},N_{i,2}=n_{i,2}] &= \Pr[N_{i,1}\leq n_{i,1},N_{i,2}\leq n_{i,2}]-\Pr[N_{i,1}\leq n_{i,1}-1,N_{i,2}\leq n_{i,2}]\\ &- \Pr[N_{i,1}\leq n_{i,1},N_{i,2}\leq n_{i,2}-1]+\Pr[N_{i,1}\leq n_{i,1}-1,N_{i,2}\leq n_{i,2}-1]\end{aligned} \tag{7.6}$$

from which we use the generalisation of Vandenhende and Lambert (2000) for T events to compute the contribution to the likelihood for any insured i as :

$$\begin{aligned}\Pr[N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}] &= \sum_{l_1=n_{i,1}-1}^{n_{i,1}}\ldots\sum_{l_T=n_{i,T}-1}^{n_{i,T}}(-1)^{\sum_j(n_{i,j}-l_j)}\Pr(N_{i,1}\leq l_1,..,N_{i,T}\leq l_T)\\ &= \sum_{l_1=n_{i,1}-1}^{n_{i,1}}\ldots\sum_{l_T=n_{i,T}-1}^{n_{i,T}}(-1)^{\sum_j(n_{i,j}-l_j)}C(F_{i,1}(l_1),...,F_{i,T}(l_T)).\end{aligned} \tag{7.7}$$

As for the other count distributions seen so far, the marginal distribution can be modeled with covariates. Additionnaly, the parameter of the copula can also be modeled with regressors, but these regressors must be time invariant since the parameter of the copula is used to model all time dependences.

This major drawback of the exchangeable dependence model is that the use of this kind of model does not allowed predictions. Indeed, even if the dependence between the number of reported claims is defined for all time periods that are in the database, we cannot suppose the kind of dependence implied for periods $T+j$, $j=1,2,....$

7.3.2 Autoregressive Dependence

As for the other models seen so far, we can suppose a decreasing dependence with the time lag separating two insured periods. Following the work of Vandenhende and Lambert (2000), this autoregressive dependence is using a first-order Markov model :

$$\begin{aligned}\Pr(N_{i,t}=n_{i,t}|n_{i,t-1},n_{i,t-2},...) &= \Pr(N_{i,t}=n_{i,t}|n_{i,t-1})\\ &= \Pr(N_{i,t}\leq n_{i,t}|n_{i,t-1})-\Pr(N_{i,t}\leq n_{i,t}-1|n_{i,t-1}),\end{aligned} \tag{7.8}$$

where the probability can be expressed as a function of a bivariate copula :

$$\Pr(N_{i,t} \leq n_{i,t} | n_{i,t-1}) = \frac{C[F_{t-1}(n_{i,t-1}), F_t(n_{i,t})] - C[F_{t-1}(n_{i,t-1}-1), F_t(n_{i,t})]}{\Pr_{t-1}(N_{i,t-1} = n_{i,t-1})}, \quad (7.9)$$

and the contribution to the likelihood of each insured i can be written as :

$$\Pr[N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}] = \Pr(N_{i,1} = n_{i,1}) \prod_{j=2}^{T} \Pr(N_{i,j} = n_{i,j} | n_{i,j-1}). \quad (7.10)$$

Covariates can also be used to model the marginal distributions, but as opposed to the exchangeable dependance model, time-varying covariates can be used to model the parameters of the copula, using appropriate transformations that respect their domains. While marginal moments can be found directly with their marginal distribution (Poisson or Negative Binomial distributions), conditional moments cannot be expressed in a closed form. Uses of equations (7.8) to (7.10) are needed to compute such moments.

7.4 Predictive Distribution

The autoregressive dependence model defined the predictive distribution directly. It only depends on the number of reported claims of the last insured period.

7.5 Numerical Application

Estimated parameters of the data applied to the copula models with Poisson and NB1 marginal distributions are exhibit in tables 7.1 and 7.2. Clearly and without surprise, we see that the copulas applied to NB1 distribution result in a loglikelihood smaller than the one obtained by the Poisson marginals.

Parameters estimates are quite the same as those observed in other models. An interesting observation is the fit exhibited by each copula since the Joe's Copula with NB1 marginal seems to exhibit a better adjustement to the data. Remember that this copula has a tail dependence at the right, meaning that an insured with high number of reported claims in a year will see an increase in his expected number of reported claims for the next year.

8. Models Comparisons

Differences between models can be analysed throught the mean and the variance of some insured profiles. The mean corresponds to the frequency part of the a priori premium, that is to say premiums for new insureds. Three profiles have been chosen :

parameter		Frank		Clayton		Joe	
λ	b0	−2.6020	(0.0428)	−2.6023	(0.0428)	−2.6127	(0.0423)
	b1	0.1015	(0.0356)	0.1014	(0.0356)	0.1037	(0.0349)
	b2	−0.2040	(0.0400)	−0.2035	(0.0400)	−0.2051	(0.0397)
	b3	−0.2648	(0.0374)	−0.2647	(0.0374)	−0.2663	(0.0369)
	b4	0.1533	(0.0350)	0.1533	(0.0351)	0.1474	(0.0346)
	b5	0.1359	(0.0308)	0.1360	(0.0308)	0.1377	(0.0304)
θ	c0	0.8298	(0.1135)	0.3639	(0.1386)	−3.0691	(0.1611)
	c5	−0.4218	(0.1392)	−0.5152	(0.1680)	−0.3716	(0.1938)
Log-Lik.		-27,006.3		-27,007.2		-27,014.1	

Tab. 7.1 – Copulas Models with Poisson Marginal

parameter		Frank		Clayton		Joe	
λ	b0	−2.6107	(0.0444)	−2.6112	(0.0444)	−2.6079	(0.0442)
	b1	0.1143	(0.0367)	0.1142	(0.0367)	0.1153	(0.0361)
	b2	−0.2074	(0.0414)	−0.2069	(0.0414)	−0.2071	(0.0411)
	b3	−0.2735	(0.0387)	−0.2732	(0.0387)	−0.2715	(0.0383)
	b4	0.1562	(0.0363)	0.1561	(0.0363)	0.1498	(0.0358)
	b5	0.1454	(0.0320)	0.1456	(0.0320)	0.1442	(0.0319)
θ	c0	0.8361	(0.1147)	0.3726	(0.1401)	−2.7734	(0.1494)
	c5	−0.4257	(0.1406)	−0.5210	(0.1698)	−0.3489	(0.1781)
α		0.1102	(0.0077)	0.1102	(0.0077)	0.1167	(0.0081)
Log-Lik.		-26,800.2		-26,801.2		-26,796.3	

Tab. 7.2 – Copulas Models with NB1 Marginal

Profile Number	Kind of Profile	v1	v2	v3	v4	v5
1	Good	0	0	1	0	0
2	Average	1	1	0	0	0
3	Bad	1	0	0	1	1

Tab. 8.1 – Profiles analyzed

Models		Low Risk		Medium Risk		High Risk	
		Mean	Variance	Mean	Variance	Mean	Variance
Time Ind	Poisson	0.0564	0.0564	0.0666	0.0666	0.1100	0.1100
	NB2	0.0563	0.0613	0.0666	0.0736	0.1102	0.1293
	NB1	0.0558	0.0619	0.0670	0.0744	0.1122	0.1245
Random Effects	MVNB	0.0565	0.0593	0.0661	0.0699	0.1077	0.1179
	P−MVNB1	0.0564	0.0593	0.0666	0.0706	0.1086	0.1092
	NB−Beta	0.0560	0.0630	0.0664	0.0753	0.1100	0.1283
AM	MPAM	0.0531	0.0531	0.0626	0.0626	0.1041	0.1041
	MNB1AM	0.0525	0.0581	0.0629	0.0696	0.1061	0.1173
INAR	Poisson	0.0565	0.0565	0.0663	0.0663	0.1088	0.1088
	NB1	0.0559	0.0623	0.0665	0.0741	0.1108	0.1235
Copula Frank	Poisson	0.0540	0.0540	0.0669	0.0669	0.1119	0.1119
	NB1	0.0530	0.0586	0.0669	0.0743	0.1114	0.1237
Copula Clayton	Poisson	0.0569	0.0569	0.0669	0.0669	0.1118	0.1118
	NB1	0.0559	0.0620	0.0669	0.0743	0.1113	0.1236
Copula Joe	Poisson	0.0562	0.0562	0.0663	0.0663	0.1108	0.1108
	NB1	0.0562	0.0627	0.0672	0.0751	0.1110	0.1239

Tab. 8.2 – A Priori Premiums

The first profile is known to be a good risk, while the last profile usually exhibits bad loss experience. Other profiles are medium risk. The Table 8.2 shows quite similar premiums for all models, where the biggest differences for the expected values are exhibited by the Artificial Marginal (AM) and the Frank's Copula models. Because the estimations of the parameters are approximately the same for all models, this result was expected since the sufficient conditions for consistency are satisfied (see (1984a,b)). Nevertheless, other elements can be analysed, such as the variance values, which is important in some forms of premium calculations and solvency analysis, that show greater variations between models.

For insured with a claim experience, the expected value or the predictive premiums are based on the number of reported claims of the last insured period, as shown in tables 8.3 and 8.4 for the medium risk profile. The time independent models are not shown since the expected value as well as other property of this distribution do not depend on past claims. As opposed to a priori premiums, we see that the choice of a model greatly influences the values of predictive premiums. Indeed, the ratio between the smallest and the largest predictive premiums is approximately 4% for a claim-free driver, goes to more than 40% for a insured who reported 1 claim in his last insured period, 108%, for 2 reported claims and attains more than 800% for a 4-claims reporter, that is to say that in this case, the largest premium is 9 times larger than the smallest one.

Since it is not based on autocorrelation of order one, as opposed to other models, the random effects models do not only depend on the number of reported claims of the last insured period but on the sum of all reported claims. For example, for a driver with an history of 10 years, the predictive premiums of those models are those shown in table 8.5. This property of the random effects model must be seen as a great advantage over the other models since the unobserved characteristics of the driver are not only shown by his last insured period, but on all his claim experience.

9. Link between Models

As shown in the preceeding sections, there are links between some of the models seen. For specific parameter restrictions, some models are nested, such as Poisson and Negative Binomial and their generalisations that account for time dependence or Poisson-based models against their Negative Binomial-based generalisation. Note that some models are non-nested to each other, such as the link between all models allowing for time dependence. Though it is not the primary purpose of this paper, many possible tests are shown in this section, to assess the model fit.

Models		Number of Reported Claims					
		0		1		2	
		Mean	Variance	Mean	Variance	Mean	Variance
Random Effects	MVNB	0.0624	0.0655	0.1174	0.1377	0.1723	0.2366
	P−MVNB1	0.0629	0.0660	0.1192	0.1407	0.1755	0.2441
	NB−Beta	0.0606	0.0713	0.1054	0.1267	0.1501	0.1843
AM	MPAM	0.0626	0.0626	0.1311	0.1311	0.2746	0.2746
	MPNB1	0.0629	0.0696	0.1330	0.1472	0.2814	0.3114
INAR	Poisson	0.0614	0.0614	0.1354	0.1299	0.2094	0.1984
	NB1	0.0618	0.0689	0.1321	0.1342	0.2023	0.1995
Copula Frank	Poisson	0.0612	0.0615	0.1489	0.1373	0.1586	0.1448
	NB1	0.0614	0.0684	0.1505	0.1559	0.1599	0.1644
Copula Clayton	Poisson	0.0612	0.0615	0.1493	0.1374	0.1557	0.1423
	NB1	0.0614	0.0685	0.1506	0.1555	0.1568	0.1611
Copula Joe	Poisson	0.0625	0.0622	0.1148	0.1135	0.2905	0.3370
	NB1	0.0618	0.0678	0.1319	0.1479	0.3117	0.4113

Tab. 8.3 – Predictive Premiums for Medium Risk (1)

Models		Number of Reported Claims 3 Mean	3 Variance	4 Mean	4 Variance
Random Effects	MVNB	0.2272	0.3748	0.2821	0.5647
	P−MVNB1	0.2318	0.3898	0.2881	0.5916
	NB−Beta	0.1949	0.2442	0.2397	0.3064
BP	Poisson	0.5751	0.5751	1.2042	1.2042
	NB1	0.5954	0.6587	1.2594	1.3933
INAR	Poisson	0.2833	0.2669	0.3573	0.3354
	NB1	0.2726	0.2648	0.3428	0.3302
Copula Frank	Poisson	0.1590	0.1451	0.1590	0.1451
	NB1	0.1607	0.1651	0.1608	0.1652
Copula Clayton	Poisson	0.1559	0.1425	0.1559	0.1425
	NB1	0.1573	0.1616	0.1574	0.1616
Copula Joe	Poisson	0.6136	0.8830	1.0737	1.7997
	NB1	0.5983	0.9606	0.9788	1.8389

Tab. 8.4 – Predictive Premiums for Medium Risk (2)

Model	A priori	Number of Reported Claims 0	1	2	3	5	10
MVNB	0.0661	0.0418	0.0785	0.1153	0.1520	0.2256	0.4093
P−MVNB1	0.0666	0.0417	0.0791	0.1165	0.1538	0.2286	0.4155
NB−Beta	0.0664	0.0432	0.0751	0.1070	0.1389	0.2028	0.3623

Tab. 8.5 – Predictive Premiums for Medium Risk (3)

9.1 Nested Models

Two situations imply nested models for the distributions seen on this paper. Firstly, to account for overdispersion of the marginal distribution, the Poisson distribution has been generalized into Negative Binomial distribution in all models allowing for time dependence. The way to test this generalisation is to check if the extra parameter accounting for the overdispersion is statistically significative. Secondly, other nested models are the time independent models and their generalisations into models where time dependence exists, like the Poisson against its generalisation into Poisson-INAR(1). As for overdispersion test, it can be done by a check on the extra parameter that implies the time dependence.

One problem with standard specification tests (Wald or Likelihood ratio tests) happens when the null hypothesis is on the boundary of the parameter space. When a parameter is bounded by the H_0 hypothesis, the estimate is also bounded and the asymptotic normality of the MLE no longer hold under H_0. Consequently, a correction must be done. Results form Chernoff (1954) for the likelihood ratio statistic and Moran (1971) for the Wald Test, reviewed by Lawless (1987) for the study of the negative binomial distribution, showed that under the null hypothesis, the LR statistic have a probability mass of $\frac{1}{2}$ on the boundary and $\frac{1}{2}\ \chi^2_{1-2\delta}$ (rather than $\chi^2_{1-\delta}$). Consequently, in this situation, one-sided test must be used. Analogous result shows that for the Wald test, where there is a mass of one half at zero and a normal distribution for the positive values. In this case, as mentionned by Cameron and Trivedi (1998), one continues to use the usual one-sided test critical value of $z_{1-\delta}$.

An other way to deal with this is to use tests that do no change properties under this kind of hypothesis, such as score tests (Moran (1971) and Chant (1974)) or Hausman (1978) test. However, testing these overdispersion situations for time dependent models or testing the parameter that implies time dependence have not been the subject of many researches until now. Consequently, in our situations, we will based our results on adapted standard tests of Wald and Likelihood Ratio tests even if we know that some improvements must be done in this area.

For more details about these tests, or other ones, refer to Cameron and Trivedi (1998) or Winkelmann (2003a) in the context of count data or Gourieroux and Monfort (1995) for a general point of view. For actuarial applications of count data when independence between contracts of the

same insured is supposed, see Boucher et al. (2006c).

9.1.1 Numerical Application

Three kinds of links can be tested by the tests of nested models :

– Overdispersion test as Poisson marginal against Negative Binomial marginal [2] ;
– Poisson against their generalized models allowing for time dependence ;
– Negative Binomial against their generalized models allowing for time dependence.

Direct application of Wald and Likelihood Ratio tests can be done on all models seen in this paper. The results are quite clear : all Poisson marginals are rejected against their overdispersed generalisation, all Poisson and Negative Binomial distributions are rejected against their generalized models allowing for time dependence. All these tests involve *p-values* less than 0.01%.

9.2 Non-nested models

Apart for the comparison between models where the marginal distribution is Poisson or Negative Binomial, the other models can be seen as non-nested models. Two kind of non-nested models exist : overlapping and strictly non-nested models. Overlapping models refer to models where neither model can be derived form the other byparameter restriction, but where some joint parameter restrictions result in the same model. Strictly non-nested models refer to models where even conjoint parameter restriction cannot produce identical model. For a general discussion on non-nested models of for more formal definitions of non-nested models are given in Pesaran (1974) and Vuong (1989).

9.2.1 Information Criteria

A standard method of comparing non-nested models refer to the information criteria, based on the fitted log-likelihood function. Since the loglikelihood increases with the addition of parameters, the criteria used to distinguish models must penalized models with large number of parameters. Developments of such methods begun with the work of Akaike (1973) and Schwartz (1978) who proposed the Akaike information Criteria $(AIC) = -2\log(L) + 2k$ and the Bayesian information Criteria $(BIC) = -2\log(L) + \log(n)k$, where k and n represent respectively the number of parameters of the model and the number of observations. Many other penalized criteria have been proposed in the statistical literature (see Kuha (2004) for an overview). However, the AIC and BIC criteria are the most often used in practice. Application of these criteria on the non-rejected models

[2] Note that the Poisson Random Effects model is not nested to the Negative Binomial distribution with random effects.

Models	LogLikelihood	AIC	BIC
MVNB	−26,690.9	53,395.78	53,462.79
Pseudo MVNB1	−26,689.8	53,393.57	53,460.58
NB1−Beta	−26,657.4	53,330.73	53,407.32
MNB1AM	−26,790.7	53,599.32	53,685.48
INAR−NB1	−26,816.0	53,649.93	53,736.09
Frank−NB1	−26,800.2	53,618.48	53,704.64
Clayton−NB1	−26,801.2	53,620.31	53,706.48
Joe−NB1	−26,796.3	53,610.61	53,696.77

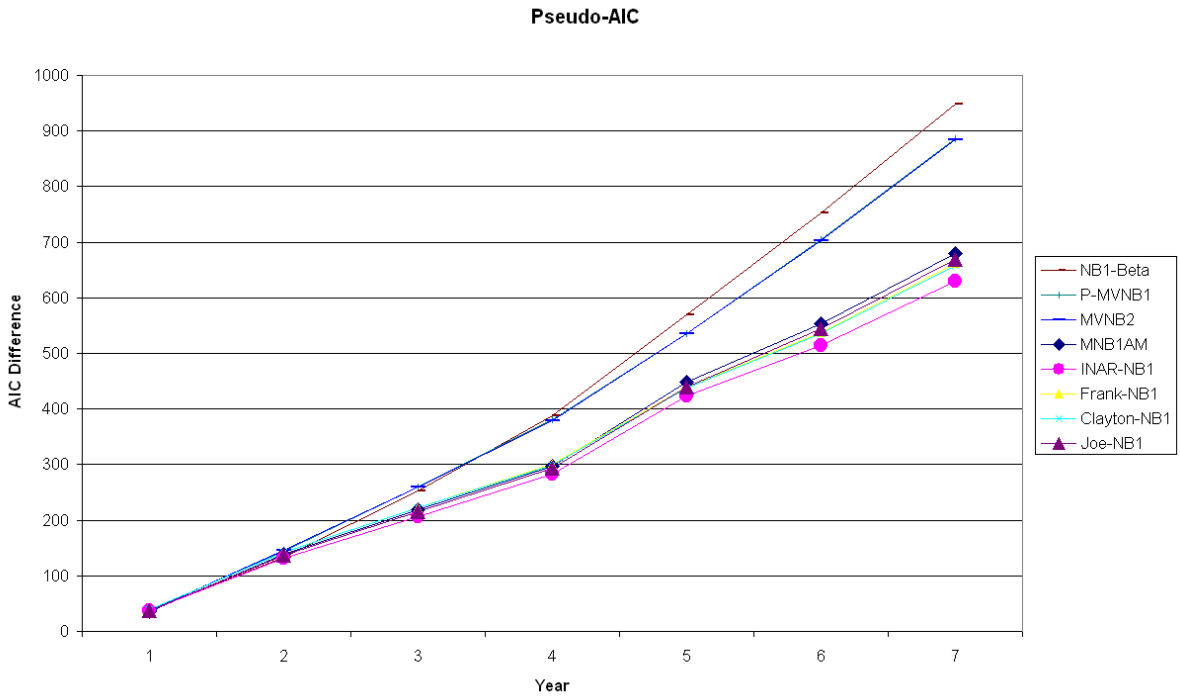
Tab. 9.1 – AIC and BIC

seen in the first section leads to the results of table 9.1.

Models do not contain the same number of parameters. Consequently, analysis of each information criteria can result in different conclusions. Note that despite their apparent simplicity, the information criteria are based on explicit theorical considerations, and AIC and BIC do not have the same foundations. As shown by Kuha (2004), refering to older papers such as Atkinson (1981) and Chow (1981), the aims of the AIC and BIC criteria are not the same. The aim of the BIC, refering to a Bayesian approach, is to identify the model with the highest probability to be true, giving that one model under investigation is true. On the other side, the AIC denies the existence of an identificable true model and, for example, minimizes the distance or discrepancy between densities. Moreover, in model selection, it has been argued that the BIC criteria penalizes large models too heavily. However, without big surprise, results of the previous table clearly show that the random effects models are better to model the insurance data than the other models seen in this paper. More particulary, we see that the NB1-Beta model is the model that provides the better fit to our data.

This analysis can be illustred differently. Indeed, we can do an intuitive analysis of the log-likelihood by models, by separating this statistic by year. To smooth the random variations between year, we analyse et illustrate the cumulative difference of fit between each remaining model and the Poisson distribution, that does not imply dependence between contracts of the same insured.

Because the models do not involve the same number of parameter, a correction factor AIC is also used. The graph below shows the values of a pseudo-AIC by year (versus the Poisson distribution) for each of the remaining models presented :



As the number of past experiences grows, we can see that the fits of the random effects models are clearly better than the Poisson distribution. Indeed, as the past experience increases, other models shows their limited predictive capacity, since the improvement in the fit stays proportionnal with the number of years since they are only based on the number of claims of the last insured period. By opposition, the random effects models seem to improve more and more their fit as the number of insured periods goes large.

9.2.2 **Vuong Test**

Random effects models clearly show a better fit than the other models. To see if the difference in the loglikelihood between the MVNB model (which shows approximately the same fit as the Pseudo MVNB1) and the NB1-Beta model is statistically significant (as illustrated in the seventh year of the last graph), a test based on the difference in the loglikelihood can be performed.

For independent observations, a loglikelihood ratio test for non-nested models, developped by

Vuong (1989) and generalised by Rivers and Vuong (2002) can be used to see if the NB1-Beta model is statistically better than the other models. This test is based on the likelihood ratio approach of Cox (1961), with a correction that corresponds to the standard deviation of this difference. As opposed to Boucher et al. (2006c) who applied directly the Vuong test on all available models, this test cannot be applied directly on our models since some observations - all contracts of the same insured - are not independent. However, as proposed by Golden (2003), this non-nested models test can be generalised quite directly to correlated observations. It can be expressed as :

$$T_{LR,NN} = \frac{\sqrt{nT}\left(\sum_i \sum_t \ell(f_{i,t},\hat{\beta}_1) - \ell(g_{i,t},\hat{\beta}_2)\right)}{\sigma_{T_{LR,NN}}}, \tag{9.1}$$

where the loglikelihood function for the distribution f is defined as :

$$\ell(f_{i,t},\hat{\beta}_1) = -\log\left[P_f(n_{i,t}|n_{i,t-1},...,n_{i,1})\right]. \tag{9.2}$$

Then, for observation at time t, conditional loglikelihood must be expressed based on experience $1,2,...,t-1$, using predictive distributions. The parameter $\sigma_{T_{LR,NN}}$ is the square root of the *discrepancy variance* and is expressed as :

$$\sigma^2_{T_{LR,NN}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{T}\left(\ell(f_{i,t},\hat{\beta}_2) - \ell(g_{i,t},\hat{\beta}_2)\right)\left(\ell(f_{i,j},\hat{\beta}_2) - \ell(g_{i,j},\hat{\beta}_2)\right). \tag{9.3}$$

The expression of the variance is closely related to the one obtained with the standard Vuong test, except that the covariances between correlated observations are considered. Note that even if the generalisation of this test for correlated observations share close similarities to the Vuong test, intermediary tests must be performed to ensure that the test is valid (Golden (2003)). First, a necessary but not sufficient condition to establish asymptotic properties of this test involves the calculation of the following matrix :

$$B = \begin{bmatrix} B_{f,f} & B_{f,g} \\ B_{g,f} & B_{g,g} \end{bmatrix}, \tag{9.4}$$

where

$$B_{f,g} = \frac{1}{n} = \sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{T}\nabla\ell(f_{i,t},\hat{\beta}_1)\nabla\ell(g_{i,j},\hat{\beta}_2). \tag{9.5}$$

The matrix B must be positive definite or, at least, must converge to it (see Golden (2003) for details). Since the variance of the difference between loglikelihood involves covariance terms, the second intermediary step is done to avoid a situation where $\sigma_{T_{LR,NN}}$ can be equal to zero. It involves the calculation of the *estimated discrepancy autocorrelation coefficient* and is defined as :

$$\hat{r} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1,j\neq t}^{T}\Big(\ell(f_{i,t},\hat{\beta}_1)-\ell(g_{i,t},\hat{\beta}_2)\Big)\Big(\ell(f_{i,j},\hat{\beta}_1)-\ell(g_{i,j},\hat{\beta}_2)\Big)}{2T\sum_{i=1}^{n}\sum_{t=1}^{T}(\ell(f_{i,t},\hat{\beta}_1)-\ell(g_{i,t},\hat{\beta}_2))^2} \tag{9.6}$$

The adapted test of Golden (2003) assumes that $\hat{r} \neq -1/(2T)$, so this assumption must be verified to ensure that the test is correctly specified. The last intermediary test to do is the one that verifies that the two models cannot be equal. Since we work with non-nested models and we already verify that models cannot be equal by simultaneous parameter restrictions, this step is not needed in our case.

To apply the adapted Vuong test for correlated observation, none of the two models has to be true. The null hypothesis of the test is that the two model are equivalent, expressed as $H_0 : E[\ell(f,\hat{\beta}_1)-\ell(g,\hat{\beta}_2)] = 0$. Under the null hypothesis, the test converge to a standard normal distribution. Rejection of the test in favor of the distribution f happens when $T_{LR,NN} > c$, or in favor of g if $T_{LR,NN} < c$, where c represents the critical value for some significance level.

Modification of this test is needed is case where the compared models do not have the same number of parameters. As proposed by Vuong (1989), we may consider the following adjusted statistic :

$$\hat{C}(\theta) = C(\theta) + K(f,g),$$

where $K(f,g)$ is a correction factor based on the difference between the information criteria (seen in the preceeding subsection) of models f and g.

For the MVNB against the NB1-Beta models, the Vuong test, adapted to correlated observations, shows that the NB1-Beta model is statistically better than the MVNB model. Indeed, the resulting test involves a *p-value* of less than 0.5% for the adapted tests (AIC and BIC). Intermediary tests show that the matrix B is positive definite, while the value of the *estimated discrepancy autocorrelation coefficient* is far from $-1/14$ since it equals 0.07.

10. Conclusion

A wide selection of models can be used to model the dependence that can exist between contracts of the same insured where each model can be interpreted differently. We showed that the choice of a model has a great impact on the predictive premiums, that is to say premiums based on a claim experience. We also showed that, despite their initial intuitive appeal, some models cannot be used. We conclude that the random effects models are the better ones to fit the data since they are more flexible to compute the next year premium, that depends not only on the past insured period but on the sum of all reported claims. More precisely, we saw that the NB1-Beta model exhibits statistically the better fit. Formally, this choice of the best distribution describing our data has been supported by specification tests for nested or non-nested models.

All the models analysed were based on generalisations of Poisson and NB1 distributions. Models seen in this paper could be applied on other claim count distributions, such as Hurdle or Zero-Inflated models who provided good fit for cross-section data (Boucher et al. (2006c)). A study in the vein of the present one could also be performed for claim severities to propose interpretations of the process that generates the dependence between the amount of claim.

Chapitre 5

Duration Dependence Models for Claim Counts

Publication prochaine dans *Blatter Deutsche Gesellschaft fur Versicherungsmathematik (German Actuarial Bulletin)*
en collaboration avec Michel Denuit

1. Introduction

There are different types of positive occurrence dependence for insured events, referring to the case where the occurrence of an event increases the probability of occurrence of the next insured event. On one hand, real positive contagion implies that the occurrence of an event increases the probability of future occurrence of the event. On the other hand, apparent positive contagion arises from the recognition of the accident proneness of an individual. Apparent positive contagion is the classical assumption in motor insurance pricing. Although past events do not truly influence the probability to report a claim, they provide some information about the "ability" of the driver. For more details, see, e.g., Pinquet (2000).

For the Negative Binomial distribution, Feller (1943) concludes to the impossibility of identifying the nature of the contagion since this distribution can result from processes implying apparent or true contagion. This can be generalized by the impossibility theorem of Bates and Neyman (1951) who stated that in a cross section of count, is impossible to distinguish between true and apparent contagion.

Duration dependence occurs when the outcome of an experiment depends on the time that has elapsed since the last success Winkelmann (2003a). Then, the occurrence of an event modifies the expected waiting time to the next occurrence of the event.

Risk classification techniques for claim counts have been the topic of many papers appeared in the actuarial literature. Usually, the starting point of the modeling of the number of reported claims is the Poisson distribution. However, this distribution suffers from some severe drawbacks that limit its use. In this paper, we build new claim counts distribution from the description of the time between two consecutive claims reported by the same policyholder. This point of view has been adopted by several papers that appeared in the actuarial literature. For instance, Lawless (1987) worked with a non-homogeneous Poisson model. More recently, Keiding et al. (1995) used the Cox model for the waiting time between claims in auto, property and household insurance.

Panel data are modelled with the help of Gamma distributed random effects. Predictive distributions are found analytically for the Exponential waiting time model, while Markov Chain Monte Carlo (MCMC) simulations are needed in the other cases. Statistical analysis shows that the models proposed in this paper substantially improve the fitting compared to the standard panel count data models.

The paper is organized as follows. The construction of the claim count distributions from renewal processes (defined by means of waiting times) is done in Section 2. Section 3 applies the resulting

distribution to the modelling of claim numbers, by taking explanatory variables and residual heterogeneity into account. Random effects are also introduced to allow for panel data. Predictive distribution generally cannot be expressed in a closed-form, but a Monte Carlo Markov Chain algorithm is built in Section 4 to obtain a sample of it. Posterior distribution based on MCMC is also presented. Section 5 concludes with a numerical example based on a Belgian motor third party liability insurance portfolio. Specification tests and comparisons between non-nested models are discussed.

2. Count Model Construction

2.1 General principle

Let τ_i be the waiting time between the the $i-1^{th}$ event and the i^{th} event. The k^{th} event thus occurs at time

$$\nu(k) = \sum_{i=1}^{k} \tau_i. \tag{2.1}$$

The τ_i's are assumed to be independent and identically distributed.

Let $N(t)$ be the number of event occurring during the interval $[0, t]$. From (2.1), we can state that the relationship between the arrival time $\tau(i)$ and the count process $N(t)$ is $\nu(k) \leq t \Leftrightarrow N(t) \geq k$. Hence,

$$\begin{aligned} P(N(t) = k) &= P(N(t) < k+1) - P(N(t) < k) \\ &= P(\nu(k+1) > t) - P(\nu(k) > t) \\ &= F_k(t) - F_{k+1}(t) \end{aligned} \tag{2.2}$$

where $F_k(t)$ is the distribution function of $\nu(k)$.

2.2 Exponential waiting times and Poisson count distribution

If the τ_i's have a common Exponential distribution, that is, if their respective probability density function is

$$f(\tau; \lambda) = \lambda e^{-\lambda\tau},$$

then we find the classical Poisson process for claim counts. This means that the number of claims occurring in the time interval $[0, t]$ is Poisson distributed with mean λt, that is,

$$P(N(t) = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

In this case, there is no duration dependence (since the Exponential distribution is known to be memoryless).

2.3 Gamma waiting times and Gamma count distribution

Winkelmann (1995) suggested to take the τ_i's Gamma distributed, with density

$$f(\tau;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\lambda\tau}, \tag{2.3}$$

where $\Gamma(\cdot)$ is the gamma function, $\alpha > 0$ and $\lambda > 0$. The Gamma hazard function is increasing for $\alpha > 1$ and is decreasing for $\alpha < 1$ (and shows constant hazard of $\alpha = 1$). Thus, for $\alpha \neq 1$, the model exhibits duration dependence since the probability to have an event depends on the time of the last event.

The Gamma distribution has mean equal α/λ and variance equal to α/λ^2. As a particular case, we find the Exponential distribution for $\alpha = 1$. The stability under convolution of the Gamma distribution for fixed λ parameter implies that $\nu(k)$ is also Gamma distributed with density

$$f(\nu;\alpha,\lambda) = \frac{\lambda^{k\alpha}}{\Gamma(k\alpha)}\nu^{k\alpha-1}e^{-\lambda\nu}.$$

Hence,

$$F_k(t) \quad = \quad \frac{1}{\Gamma(k\alpha)}\int_0^t \lambda^{k\alpha}\nu^{k\alpha-1}e^{-\lambda\nu}d\nu = G(\alpha k,\lambda t)$$

where the integral is known as an incomplete gamma function. The Gamma count distribution (GCD) can be found using equation (2.2) :

$$P(N(t)=k) = G(\alpha k,\lambda t) - G(\alpha k+\alpha,\lambda t)$$

with $G(0,\alpha\lambda) = 1$. This probability can be calculated using the probgam function in the SAS system. Additionally, the incomplete gamma function can be evaluated using integrations approximations or asymptotic expansions Abramowitz and Stegun (1968).

2.4 Weibull waiting times and Weibull count distributions

Another common model used in the duration analysis is the Weibull distribution. In this case, the τ_i's have density

$$f(\tau;c,\lambda) = \lambda c\tau^{c-1}\exp(-\lambda\tau^c)$$

for $\lambda > 0$ and $c > 0$. The Exponential distribution corresponds to $c = 1$. The Weibull hazard function is monotonically increasing for $c > 1$ and monotonically decreasing for $c < 1$. The Weibull distribution has mean $\lambda^{-1/c}\Gamma(1 + 1/c)$ and variance $\lambda^{-2/c}\big(\Gamma(1 + 2/c) - \Gamma^2(1 + 1/c)\big)$.

The Weibull distribution is not stable under convolution. Nevertheless, to find the time arrival of the k^{th} event, Bradlow et al. (2006) use the k-fold convolution of the interarrival time distribution with the help of a Taylor series approximation.

This gives the following expression for the Weibull count distribution (WCD) :

$$P(N(t) = k) = \sum_{j=k}^{\infty} \frac{(-1)^{j+k}(\lambda t^c)^j \alpha_j^p}{\Gamma(cj + 1)} \tag{2.4}$$

where

$$\begin{aligned} \alpha_j^0 &= \frac{\Gamma(cj + 1)}{\Gamma(j + 1)} \\ \alpha_j^{p+1} &= \sum_{m=p}^{j-1} \alpha_m^p \frac{\Gamma(cj - cm + 1)}{\Gamma(j - m + 1)}, \quad p = 0, 1, 2, ... \quad j = p + 1, p + 2, ... \end{aligned}$$

Note that when $c = 1$, we get the Poisson distribution.

3. Claim Count Distributions

3.1 Definition

At this stage, we have three distributions for the number of claims N generated by a policyholder who has been covered by an insurance company for a period of length t : the classical Poisson distribution, for which

$$P(N = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!},$$

where λ is the annual expected claim frequency, the Gamma count distribution for which

$$P(N = k) = e^{-\lambda t} \sum_{j=0}^{\alpha-1} \frac{(\lambda t)^{\alpha k+j}}{(\alpha k + j)!},$$

and the Weibull count distribution for which

$$P(N = k) = \sum_{j=k}^{\infty} \frac{(-1)^{j+k}(\lambda t^c)^j \alpha_j^p}{\Gamma(cj + 1)}.$$

Having observed the number of claims reported by a portfolio comprising n insured drivers, we can build a regression model to explain λ in function of the observable characteristics of each of them (summarized in a vector x of explanatory variables), taking the risk exposure t into account.

As opposed to the Exponential and the Gamma interarrival time models, an interesting property of the Weibull distribution is the modeling of the time length. Indeed, for Exponential distribution, the expected claim frequency is proportional to the exposure-to-risk t while in the Weibull case t^c is used instead. In standard count modeling, the mean parameter is defined as $\lambda = \exp(x\beta + \ln(t))$, where $\ln(t)$ is known as an offset. Using the Weibull modeling, the parameter can be set to be equal to $\lambda = \exp(x'\beta + c\ln(t))$. This modeling shares similarity with a model proposed by Winkelmann (2003a), who uses a Poisson regression model where the coefficient of $\ln(t)$ as a unknown parameter.

3.2 Heterogeneity

3.2.1 Poisson distribution

Insurance data often exhibits overdispersion that may be caused by the missing of some important classification variables (swiftness of reflexes, aggressiveness behind the wheel, consumption of drugs, etc.). Equidispersion implied by the Poisson distribution is usually corrected by the introduction of a random heterogeneity term θ of mean 1 and variance A Boyer et al. (1992), and Lemaire (1995). Specifically, let N_i be the number of claims reported by policyholder i, with observable characteristics x_i and risk exposure t_i. We assume that given θ, N_i is Poisson distributed with mean $t_i\lambda_i\theta$, $i = 1, ..., n$, where the mean parameter $\lambda_i = \exp(x_i'\beta)$ includes explanatory variables x_i. An easy way of handling the heterogeneity is to introduce a Gamma distribed random effect, with both parameters equal to $1/A$. This yields a Negative Binomial distribution for observed claim counts. Winkelmann and Zimmermann (1991) and Winkelmann and Zimmermann (1995) proposed a generalization of this distribution, called "Generalized Event Count" ($GECk$), assuming that θ_i is Gamma distributed with both parameters equal to $(t_i\lambda_i)^{1-k}/A$. The $GECk$ can handle overdispersion as well as underdispersion. If we restrict ourselves to the overdispersion case only, the $GECk$ probability mass function is

$$P(N_i = n_i) = (1 + A(t_i\lambda_i)^k)^{-(t_i\lambda_i)^{1-k}/A} \prod_{j=1}^{n_i} \left(\frac{\lambda_i + A(j-1)(t_i\lambda_i)^k}{(1 + A(t_i\lambda_i)^k)j} \right)$$

where the $\lambda_i = \exp(x_i'\beta)$. This model is interesting because it covers the NB2 distribution (obtained with $k = 1$) and the NB1 distribution (obtained for $k = 0$). Its variance is equal to $t_i\lambda_i + A(t_i\lambda_i)^{k+1}$. These models reduce to the Poisson distribution in the case where $A \to 0$.

3.2.2 Gamma count distribution

To account for individual variations between interarrival rates, a Gamma distributed heterogeneity factor is superposed to the λ parameter of the Gamma distribution. Denoting as $h(\cdot)$ the

Gamma density with both parameters equal to $1/A$, we get

$$\begin{aligned} P(N_i = n_i) &= \int_0^\infty P(N_i = n_i|\theta_i)h(\theta_i)d\theta_i \\ &= \int_0^\infty \Big(G(\alpha n_i, t_i\lambda_i\theta_i) - G(\alpha n_i + \alpha, t_i\lambda_i\theta_i)\Big)h(\theta_i)d\theta_i, \end{aligned}$$

where

$$\begin{aligned} \int_0^\infty G(\alpha n_i, t_i\lambda_i\theta_i)h(\theta_i)d\theta_i &= \int_0^\infty \int_0^{t_i} \frac{(t_i\lambda_i\theta_i)^{n_i\alpha}}{\Gamma(n_i\alpha)}\nu^{k\alpha-1}e^{-t_i\lambda_i\theta\nu}\frac{(1/A)^{1/A}}{\Gamma(1/A)}\theta_i^{1/A-1}e^{-\theta/A}d\nu d\theta_i \\ &= \int_0^{t_i} \frac{\lambda_i^{n_i\alpha}}{\Gamma(k\alpha)\Gamma(1/A)}(1/A)^{1/A}\nu^{n_i\alpha-1}\left(\int_0^\infty \theta_i^{1/A+n_i\alpha-1}e^{-\theta_i(\lambda_i\nu+1/A)}d\theta_i\right)d\nu \\ &= \int_0^{t_i} \frac{\Gamma(1/A+n_i\alpha)}{\Gamma(n_i\alpha)\Gamma(1/A)}\left(\frac{1/A}{\lambda_i\nu+1/A}\right)^{1/A}\left(\frac{\lambda_i\nu}{\lambda_i\nu+1/A}\right)^{n_i\alpha}\nu^{-1}d\nu \end{aligned}$$

or using asymptotic expansions :

$$\begin{aligned} \int_0^\infty G(\alpha n_i, t_i\lambda_i\theta_i)h(\theta_i)d\theta_i &= \int_0^\infty \frac{1}{\Gamma(n_i\alpha)}\sum_{i=0}^\infty \frac{(-1)^i(t_i\lambda_i\theta_i)^{n_i\alpha+i}}{(n_i\alpha+i)i!}\frac{(1/A)^{1/A}}{\Gamma(1/A)}\theta_i^{1/A-1}e^{-\theta/A}d\theta_i \\ &= \frac{(1/A)^{1/A}}{\Gamma(k\alpha)\Gamma(1/A)}\sum_{i=0}^\infty \frac{(-1)^i(t_i\lambda_i)^{n_i\alpha+i}}{(n_i\alpha+i)i!}\left(\int_0^\infty \theta_i^{1/A+n_i\alpha+i-1}e^{-\theta_i/A}d\theta_i\right) \\ &= \frac{1}{\Gamma(n_i\alpha)\Gamma(1/A)}\sum_{i=0}^\infty \frac{(-1)^i(At_i\lambda_i)^{n_i\alpha+i}}{(n_i\alpha+i)i!}\Gamma(1/A+n_i\alpha+i) \end{aligned}$$

If we set the parameter α to 1, this heterogeneous model becomes a NB2 distribution. If, instead of using a Gamma distribution with both parameters equal to $1/A$, we choose to use a Gamma distribution with parameters λ/A, the model is a generalization of a NB1 distribution. All these random effects models come back to the GCD in the situation where $A \to 0$. Expected value and variance have to be computed numerically.

3.2.3 Weibull count distribution

To model the individual specificities, a Gamma distributed heterogeneity factor can be superposed to the λ_i parameter of the WCD, giving

$$\begin{aligned} P(N_i = n_i) &= \int_0^\infty P(N_i = n_i|\theta_i)h(\theta_i)d\theta_i \\ &= \int_0^\infty \left(\sum_{j=n_i}^\infty \frac{(-1)^{j+n_i}(\lambda_i \theta_i t_i^c)^j \alpha_j^p}{\Gamma(cj+1)} \right) \frac{(1/A)^{1/A}}{\Gamma(1/A)} \theta_i^{1/A-1} e^{-\theta_i/A} d\theta_i \\ &= \frac{(1/A)^{1/A}}{\Gamma(1/A)} \left(\sum_{j=n_i}^\infty \frac{(-1)^{j+n_i}(\lambda t_i^c)^j \alpha_j^p}{\Gamma(cj+1)} \int_0^\infty \theta_i^{1/A+j-1} e^{-\theta_i/A} d\theta_i \right) \\ &= \sum_{j=n_i}^\infty \frac{(-1)^{j+n_i}(A\lambda_i t_i^c)^j \alpha_j^p \Gamma(1/A+j)}{\Gamma(cj+1)\Gamma(1/A)} \end{aligned}$$

When $c = 1$, the WCD with Gamma heterogeneity becomes the NB2 distribution. Changing $1/A$ into λ/A leads the heterogeneity model to become a generalization of a NB1 distribution. All these random effects models come back to the WCD in the situation where $A \to 0$. As for the GCD, the first moments of the model must be evaluated numerically.

3.3 Allowance for panel data

3.3.1 General principle

Insurance data originate from a repetition of one-year contracts for an insured, where some dependence over contracts of the same insured exists. Longitudinal data (or panel data) consists of repeated observations of individual units that are observed over time. Each individual is assumed to be independent but correlation between observations relating to the same individual is permitted. This fact must be regarded as an advantage and must be used in the construction of frequency models.

The random effect term, having mean 1 and variance A, captures hidden features of the risk. Given the insured-specific heterogeneity term, θ_i, the annual claim numbers $N_{i,1}, N_{i,2}, \ldots, N_{i,T}$ are independent. The joint probability function of $N_{i,1}, ..., N_{i,T}$ is thus given by

$$\begin{aligned} P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) &= \int_0^\infty P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}|\theta_i)h(\theta_i)d\theta_i \\ &= \int_0^\infty \left(\prod_{\tau=1}^T P(N_{i,\tau} = n_{i,t}|\theta_i) \right) h(\theta_i)d\theta_i \end{aligned}$$

3.3.2 Poisson distribution

The joint distribution in a closed form. Indeed, the joint distribution of the number of claims, when the random effects follow a Gamma distribution of mean 1 and variance A, is equal to

$$P(N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T}) = \left(\prod_{\tau=1}^{T}\frac{(t_{i,\tau}\lambda_{i,\tau})^{n_{i,\tau}}}{n_{i,\tau}!}\right)\frac{\Gamma(\sum_{\tau=1}^{T}n_{i,\tau}+1/A)}{\Gamma(1/A)}\left(\frac{1/A}{\sum_{\tau=1}^{T}t_{i,\tau}\lambda_{i,\tau}+1/A}\right)^{1/\alpha}\left(\sum_{\tau=1}^{T}t_{i,\tau}\lambda_{i,\tau}+1/A\right)^{-\sum_{\tau=1}^{T}n_{i,\tau}}.$$

This distribution is known as a Multivariate Negative Binomial or Negative Multinomial (MVNB). Note that this distribution can also be seen as the generalization of the bivariate Negative Binomial of Marshall and Olkin (1990). For this distribution, $E[N_{i,\tau}] = t_{i,\tau}\lambda_{i,\tau}$ and $Var[N_{i,\tau}] = t_{i,\tau}\lambda_{i,\tau} + A(t_{i,\tau}\lambda_{i,\tau})^2$.

3.3.3 Gamma count distribution

Panel data model for GCD is constructed as above. A generalization of the MVNB distribution can be found if random effects θ_i are supposed to follow a Gamma distribution :

$$\begin{aligned}&P(N_{i,1}=n_{i,1},...,N_{i,T}=n_{i,T})\\&=\int_0^\infty\prod_{\tau=1}^{T}\Big(G(\alpha n_{i,\tau},t_{i,\tau}\lambda_{i,\tau}\theta_i)-G(\alpha n_{i,\tau}+\alpha,t_{i,\tau}\lambda_{i,\tau}\theta_i)\Big)h(\theta_i)d\theta_i.\end{aligned}$$

Complex computations are needed to obtain closed form from this last equation of the Multivariate Gamma Count Distribution (MVGCD). Here, we use the NLMIXED procedure of SAS, as described by Nelson et al. (2006). To approximate densities involving random effects, the NLMIXED procedure uses numerical evaluation techniques such as Gaussian quadrature. This technique is only available for Normally distributed random effects. Nevertheless, non-Normal random effects can be accommodated within the numerical integration techniques via the use of the probability integral transformation. Since the NLMIXED procedure allows for Normal random effects ϵ, application of the probability integral transform to the normal random effect such as $u = \Phi(\epsilon)$ leads to $b = F^{-1}(u)$ which has density $f(b)$. As for the GCD model with heterogeneity, the moments can be estimated numerically.

3.3.4 Weibull count distribution

The joint distribution for all contracts of the same insured is expressed as the one seen with the panel data model of the GCD :

$$\begin{aligned} & P(N_{i,1} = n_{i,1}, ..., N_{i,T} = n_{i,T}) \\ & = \int_0^\infty \prod_{\tau=1}^{T} \left(\sum_{j=n_{i,\tau}}^{\infty} \frac{(-1)^{j+n_{i,\tau}} (\lambda_{i,\tau} \theta_i t_{i,\tau}^c)^j \alpha_j^p}{\Gamma(cj+1)} \right) h(\theta_i) d\theta_i. \end{aligned}$$

This distribution is called the Multivariate Weibull Count Distribution (MVGCD). Parameters can be estimated by maximum likelihood using the NLMIXED procedure in SAS.

4. Predictive Distribution

4.1 Definition

The predictive distributions of panel data with random effects depend on the past experience of each insured. Indeed, at each insured period, the heterogeneity term θ_i can be updated based on past claims experience. Formally, the predictive distribution at time $T+1$ can be computed as :

$$\begin{aligned} P(N_{i,T+1} = n_{i,T+1} | n_{i,1}, ..., n_{i,T}) &= \int P(N_{i,T+1} = n_{i,T+1} | \theta_i) \left(\frac{\left(\prod_\tau P(N_{i,\tau} = n_{i,\tau} | \theta_i)\right) h(\theta_i)}{\int \left(\prod_\tau P(N_{i,\tau} = n_{i,\tau} | \theta_i)\right) h(\theta_i) d\theta_i} \right) d\theta_i \\ &= \int P(N_{i,T+1} = n_{i,T+1} | \theta_i) g(\theta_i | n_{i,1}, ..., n_{i,T}) d\theta_i \end{aligned}$$

where $g(\theta_i | n_{i,1}, ..., n_{i,T})$ is called the a posteriori distribution of the random effects θ_i, reflecting the past claims experience of insured i. If this a posteriori distribution can be expressed in closed form, moments of the predictive distribution can be found easily be conditioning on the random effects θ_i.

4.2 Poisson distribution

As it is well known, we found that the a posteriori distribution of the heterogeneity term for the Poisson model with Gamma random effects is also a Gamma distributed with parameters $\sum_\tau^T t_{i,\tau} \lambda_{i,\tau} + 1/A$ and $\sum_\tau^T n_{i,\tau} + 1/A$. The a posteriori expected claim frequency is equal to :

$$E[N_{i,T+1} | n_{i,1}, ..., n_{i,T}] = \lambda_{T+1} \frac{\sum_\tau^T n_{i,\tau} + 1/A}{\sum_\tau^T t_{i,\tau} \lambda_{i,\tau} + 1/A}.$$

4.3 Gamma count distribution

In this case, exact predictive and posterior distributions for the random effects can only be evaluated using numerical computations or simulations. Here, we use Markov Chain Monte Carlo (MCMC) simulations to compute posterior and predictive distributions.

The idea of MCMC simulations is to simulate realizations from a Markov chain whose stationary distribution is the joint distribution of the random effects (Smith and Roberts (1993), for example). We refer the reader, e.g., to Scollnik (2001) for an introduction to MCMC simulations in actuarial sciences. To obtain an approximation of the a posteriori distribution of the random effect, we use the Gibbs sampler algorithm with the BUGS (Bayesian inference Using Gibbs Sampler) software package (see Scollnik (2001) for an overview of BUGS and its applications in actuarial science).

Once the discarding and the thinning has been done, the monitoring of the convergence of the n simulated values of each of the m chains can be done. Graphical analysis comparing each simulated chains is a common alternative. Checking the expected value, the variance and the median of each simulated chain can also be done. An other way to monitor the convergence is to compute the following statistics :

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{\psi}_{.j} - \bar{\psi}_{..})^2, \quad W = \frac{1}{n}\sum_{j=1}^{m} s_j^2$$

where

$$\bar{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij}, \quad \bar{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\psi_{.j} \text{ and } s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \bar{\psi}_{.j})^2.$$

The estimated statistics B and W can be seen as the between and the within-sequence variances. We can estimate the variance of the marginal posterior distribution by a weighted average of B and W, such as :

$$\hat{Var}(\psi) = \frac{n-1}{n}W + \frac{1}{n}B$$

The estimated values W and $\hat{Var}(\psi)$ can be seen respectively as the bottom and the upper bound of the real variance of ψ. In the limit, both W and $\hat{Var}(\psi)$ approach $Var(\psi)$, but from opposite directions. A ratio between the upper and the lower bound of the variance can be computed, which will estimate the factor by which the upper bound might be reduced. This ratio is called the estimated potential scale reduction :

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{Var}(\psi)}{W}} \tag{4.1}$$

If $\hat{R}$ is near 1 (less than 1.1 seems to be generally accepted) for all scalar estimated of interest, the convergence is accepted and the chain of simulations can be seen as the needed posteriori distribution.

4.4 Weibull count distribution

Posterior distribution of the random effects can be approximate using MCMC simulations, as expressed in the GCD section. As for the Gamma Waiting Time Count Distribution, random effects models come back to the WCD in the situation where $A \to 0$ and the first moments of the model must be evaluated numerically.

5. Numerical Application

In this paper, we work with a Belgian motor third party liability insurance portfolio comprising 45,350 observations, which are policies followed for up to 3 consecutive years (from 1997 to 1999). The data exhibits a frequency of claims of 18.4%, while the maximum number of reported claims is 5. Approximately half (51.2%) of the contracts exhibits a entire period of coverage, while the average covered period is 0.78. For each contract, we have informations about the annual number of claims together with some characteristics of the insured :

- Variable $b1$: equals 1 if the driver is a woman ;
- Variable $b2$: equals 1 if the age of the driver is less than 22 years old ;
- Variable $b3$: equals 1 if the age of the driver is between 23 and 30 years old ;
- Variable $b4$: equals 1 if the size of the city where the insured was living is big (based on the number of residents) ;
- Variable $b5$: equals 1 if the size of the city where the insured was living is small.

As stated earlier, estimation of the parameters is performed by maximum likelihood. We use the NLMIXED procedure in SAS to find the estimates and their corresponding standard error numerically. Many starting values have been used to ensure that convergence does not attain local maximum.

5.1 Parameters estimation

When the models are used to fit cross-section data, that it is to say when all observations are independent form each other, one can argue that heterogeneity on GCD and the WCD is not necessary. Indeed, the GCD and the WCD imply overdispersion which often removes the need of an heterogeneity term. However, for panel data, the random effects remain useful since they account for the correlation between observations of the same insured.

We use Gamma random effects distributions having mean 1 and variance A to fit our data. When applied to our insurance data, it leads to the following results :

Parameter	MVNB		MVGCD		MVWCD		MVNB*	
b0	−1.8890	(0.0296)	−6.4378	(0.4852)	−2.0711	(0.0286)	−2.0343	(0.0297)
b1	0.6564	(0.0639)	1.4864	(0.1822)	0.5540	(0.0590)	0.5737	(0.0623)
b2	0.2520	(0.0281)	0.6169	(0.0824)	0.2268	(0.0261)	0.2399	(0.0275)
b3	−0.0587	(0.0287)	−0.1340	(0.0723)	−0.0474	(0.0264)	−0.0486	(0.0280)
b4	0.2603	(0.0338)	0.6360	(0.0949)	0.2339	(0.0311)	0.2467	(0.0329)
b5	0.0735	(0.0340)	0.1843	(0.0869)	0.0674	(0.0315)	0.0680	(0.0332)
d	.	.	.	.	.	.	0.3219	(0.0237)
α, c	.	.	0.3413	(0.0243)	0.3682	(0.0239)	.	.
A	1.9861	(0.1500)	2.0637	(0.7308)	12.8707	(1.7241)	2.5109	(0.2447)
Loglike.	-19,649.3		-19,400.6		-19,398.4		-19,350.7	

Non-constant hazard models improve the fit compared to the MVNB distribution that does not imply duration dependence. Analysis of the parameters α and c implies that the data exhibits positive duration contagion. Consequently, within each year of contract, the report of a claim decreases the expected time to report an other claim. We add the model MVNB* that consists of a MVNB model with an additional regressor d on $\ln(time)$. It is interesting to see that the introduction of this covariate improves a lot the fitting of the MVNB model.

5.1.1 Parameters Comparison

The parameters values cannot be compared directly within models. Indeed, regression is not done on the count data, but on the waiting times. Consequently, as explained by Winkelmann (1995), partial effects of the models must be found to make estimated parameters comparable. The

strategy is to hold all explanatory variables constant at their mean and to compute $g = \Delta\hat{N}/\Delta x$. In this situation, x is the remaining explanatory variable and the change is defined as a change of 0 to 1 for the dummy variables (or by a unit increase in the mean value for continuous variables). This measures the estimated effects of this explanatory variable on the mean, when all the other variables are constant. In the following table, the values of comparable parameters have been calculated for the MVNB, the MVGCD, the MVWCD and the MVNB* model :

Parameter	MVNB	MVGCD	MVWCD	MVNB*
b1	0.168	0.129	0.128	0.120
b2	0.048	0.040	0.041	0.039
b3	−0.011	−0.009	−0.008	−0.008
b4	0.050	0.042	0.042	0.041
b5	0.014	0.012	0.011	0.011

The impacts on the frequency of claims of the covariates are not exactly the same depending on the models. One of the major differences of the models come from the modeling of insureds having less than one complete year of experience. Young drivers exhibit many more partial insured periods (the mean of our portfolio for these insureds are 0.7 while other insureds exhibits mean of 0.78). The impact of the age is reduced by model construction. Nevertheless, we can see that the MVNB* model seems to have comparable parameter values with the MVGCD and the MVWCD models.

5.2 Expected claim frequency analysis

5.2.1 A priori expected claim frequencies

The difference between models can be shown by the calculation of the a priori expected claim frequencies. The expected number of claim and its corresponding variance are calculated for the insured profiles described in the next table :

Profiles analyzed

Profile Number	Kind of Profile	v1	v2	v3	v4	v5
1	Good	0	0	1	0	0
2	Average	0	1	1	0	1
3	Bad	1	0	0	1	0

The first profile is classified as a good driver, while the last profile exhibits bad loss experience. The other profile is medium risk. Frequency part of the premium for new insureds can be computed by this table :

Models	1^{st} Profile Mean	1^{st} Profile Variance	2^{nd} Profile Mean	2^{nd} Profile Variance	3^{rd} Profile Mean	3^{rd} Profile Variance
MVNB	0.1426	0.1830	0.1975	0.2749	0.3782	0.6623
MVGCD	0.1256	0.1391	0.1713	0.1960	0.3121	0.3925
MVWCD	0.1260	0.1394	0.1718	0.1967	0.3087	0.3893
MVNB*	0.1246	0.1635	0.1645	0.2416	0.2970	0.5185

We can see big differences between the a priori premiums, depending of the model used. The MVNB model shows premiums that are higher than the other distributions. This is also caused by the modeling of the exposure time since other models imply higher premiums for insureds having partial exposure. To understand correctly this characteristic of the model, we compute the resulting premiums for the second profile :

Time	MVNB		MVGCD		MVWCD		MVNB*	
0.9	0.1777	(0.2404)	0.1643	(0.1871)	0.1650	(0.1883)	0.1638	(0.2312)
0.8	0.1580	(0.2075)	0.1571	(0.1780)	0.1577	(0.1791)	0.1577	(0.2202)
0.7	0.1382	(0.1762)	0.1491	(0.1679)	0.1491	(0.1680)	0.1511	(0.2084)
0.6	0.1185	(0.1463)	0.1404	(0.1571)	0.1407	(0.1576)	0.1438	(0.1957)
0.5	0.0987	(0.1181)	0.1309	(0.1454)	0.1311	(0.1455)	0.1356	(0.1817)
0.4	0.0790	(0.0914)	0.1205	(0.1327)	0.1205	(0.1329)	0.1262	(0.1662)
0.3	0.0592	(0.0662)	0.1079	(0.1178)	0.1080	(0.1179)	0.1150	(0.1482)
0.2	0.0395	(0.0426)	0.0927	(0.1001)	0.0922	(0.0995)	0.1010	(0.1265)
0.1	0.0197	(0.0205)	0.0719	(0.0764)	0.0713	(0.0757)	0.0808	(0.0971)

As shown in the table above, new insureds, that are usually insured in the middle of the insured period, exhibit big different premiums depending on the model. Even if the MVNB* distribution does not suppose a clear waiting time distribution, we see that it can be used to copy GCD or WCD since its premiums are quite the same as the ones obtained by these models.

5.2.2 **Predictive Distribution**

An other comparison that can be done is the analysis of the predictive distribution. We choose again the second profile and select specific pattern of claim reporting to see the impact on the predictive premium. For the MVNB and MVNB* distributions, analytical results were available while MCMC simulations must be done for the other models. The results from 5 chains of $20,000$

simulations using the Gibbs sampler of BUGS are shown in the next table. The first 10,000 iterations of each chain have been discarded and we choose a lag of 5 to remove serial correlation. Values of R form the equation (4.1) were varying between 0.999 and 1.001, revealing the convergence of the MCMC simulations.

		MVNB		MVGCD		MVWCD		MVNB*	
T	$\sum_t n_{i,t}$	Premium	Variance	Premium	Variance	Premium	Variance	Premium	Variance
A Priori		0.1975	0.2749	0.1713	0.1960	0.1718	0.1967	0.1695	0.2416
1	0	0.1418	0.1818	0.1686	0.1926	0.1696	0.1941	0.1189	0.1544
	1	0.4235	0.5428	0.1814	0.2084	0.1831	0.2114	0.4174	0.5420
	2	0.7052	0.9039	0.1923	0.2221	0.1950	0.2266	0.7159	0.9296
2	0	0.1107	0.1350	0.1660	0.1892	0.1671	0.1912	0.0916	0.1126
	1	0.3304	0.4031	0.1792	0.2054	0.1800	0.2070	0.3214	0.3953
	2	0.5502	0.6712	0.1900	0.2192	0.1924	0.2230	0.5513	0.6781
5	0	0.0667	0.0755	0.1612	0.1835	0.1605	0.1823	0.0542	0.0616
	1	0.1991	0.2255	0.1734	0.1982	0.1736	0.1992	0.1902	0.2161
	2	0.3316	0.3755	0.1846	0.2124	0.1860	0.2148	0.3263	0.3707
	5	0.7289	0.8255	0.2149	0.2506	0.2234	0.2637	0.7345	0.8344
10	0	0.0401	0.0433	0.1507	0.1703	0.1510	0.1704	0.0322	0.0349
	1	0.1198	0.1293	0.1641	0.1867	0.1630	0.1854	0.1132	0.1224
	2	0.1995	0.2154	0.1753	0.2005	0.1752	0.2005	0.1942	0.2099
	5	0.4385	0.4735	0.2042	0.2368	0.2116	0.2476	0.4371	0.4725
	10	0.8369	0.9036	0.2470	0.2925	0.2697	0.3271	0.8420	0.9101

Contrarily to a priori expected claim frequencies, we can see that posterior results exhibit great differences between MVNB* and the MVGCD and MVWCD models. Gamma and Weibull waiting time model show similar results. However, even the MVNB* distribution is widely different from the MVGCD and the MVWCD. These differences are caused by the random effects : it shows that the good fit of the MVNB models comes from the random effects added to the distributions. By opposition, we can say that the construction of these two models seems to be better suited for the number of reported claims since they do not rely too much on the random effects to provide a good fit.

5.3 Models Comparison

As shown in the introduction of the GCD and the WCD, both models come back to the Poisson distribution for a specific value of their extra parameter. By analogy, the MVGCD and the MVWCD can also be tested to evaluate their differences with the MVNB model. This can be tested using standard Wald or Likelihood Ratio tests. Direct computation of the *p-value* clearly leads to a rejection (≤ 0.001) of the MVNB against its generalizations.

Another situation implies nested models where distributions Poisson, GCD and WCD are generalized to allow for time dependence. The way to test this generalization is to check if the extra parameter that accounts for time dependence is statistically significant. However, a problem with standard specification tests (Wald or Likelihood ratio tests) happens in this situation since the null hypothesis is on the boundary of the parameter space. Using appropriate testing procedures that can be found, e.g., in Cameron and Trivedi (1998), all extra parameters are shown to be statistically significant (*p-value* ≤ 0.01), which leads to the rejection of all 4 models implying independent contracts for the same insured.

Remaining models (MVGCD, MVWCD and MVNB*) cannot be compared directly since they are non-nested models. Two kind of non-nested models exist : overlapping and strictly non-nested models. Overlapping models refer to models where neither model can be derived from the other parameter restriction, but where some joint parameter restrictions result in the same model. A standard method of comparing non-nested models refer to the information criteria, such as the Akaike information Criteria ($AIC = -2\log(L) + 2k$ where k represent the number of parameters of the model). Because all remaining models have the same number of parameters, AIC will lead to the same conclusion as an analysis of the loglikelihood. A rule-of-the-thumb says that a difference greater than 10 indicates a significant difference between models (Burnham and Anderson (2002)). Consequently, the MVNB* seems to be the better model to fit our data since the difference between its AIC and the other ones is greater than 95.

Formally, for independent observations, a loglikelihood ratio test for non-nested models, developed by Vuong (1989) and generalized by Rivers and Vuong (2002) can be used. This test is based on the likelihood ratio approach, with a correction that corresponds to the standard deviation of this difference. As proposed by Golden (2003), this non-nested models test can be generalized

quite directly to correlated observations. Because complex intermediary steps are needed to use this statistical test (gradient evaluation, autocorrelation check, etc.), we favor AIC analysis in this paper.

6. Conclusion

Models presented in this paper suggest that duration dependence may be present in insurance data. We worked with third party liability, meaning that when an accident occurs, it shortens the expected time for another claim to be reported during the same insured period. We showed that the adapted MVNB model having a regressor on the $\ln(time)$ variable is the best model to fit our data. However, we also saw that this model rely too much on the random effects since it shows great discrepancy of premiums for some specific pattern of claim reports. We can suppose that such models can even better model the statistics of stolen cars since it is well known that this kind of insurance shows great duration contagion.